In This Issue

Charles Stelzried and Michael Klein

The name of the IPN-ISD Technology and Science News has been changed to the IND Technology and Science News. This is consistent with our recent JPL directorate name change from IPN-ISD (Interplanetary Network and Information Systems Directorate) to the Interplanetary Network Directorate (IND). The following provides an overview of the articles. These are separated into four sections: Technology–Communications Systems, Technology–Information Systems, Technology–non Deep Space (a new category initiated with this issue), and Science.

Technology—Communications Systems

Fabrizio Pollara and Kenneth Andrews (Deep Space Network Turbo Decoder Infu-

sion) describe the enhanced performance and reduced Turbo decoder complexity for the DSN (see issue 11, January 2000, for a previous discussion of Turbo Codes in the DSN). These decoders will support missions launching during and after 2003. Present Turbo code designs provide a 0.9 decibel (dB) improvement in link performance over the current best codes used by deep space missions. In addition, the lower decoder complexity reduces implementation, operations and maintenance costs.

Abhijit Biswas (Large-Aperture Antenna for Deep Space Optical Communications) provides a clear description of the future of deep space communications at optical wavelengths. Early optical systems will consist of ground-based infrastructures with large (10-meter) aperture antennas. These have a spherical primary and a field of view large enough to accommodate the blind pointing accuracy, yet small enough to reduce the background light. This design provides an exciting preview of optical ground systems in the future DSN.

Inam Haque (Ka-Band Traveling Wave Tube Amplifier) has been successfully managing the development by Thales Electronique of a 35-watt, high-efficiency, Ka-band, Traveling Wave Tube Amplifier (TWTA) proto-flight model scheduled for delivery mid-2002. This is a key microwave component for the future use of Ka-band for spacecraft to Earth (via DSN) improved communication performance. This TWTA will be used for the 2005 Mars Reconnaissance Orbiter mission back-up communications system.

C O N T E N T S

Technology—Information Systems

Aaron Kiely, Matthew Klimesh, and Justin Maki (ICER on Mars: Wavelet-Based Image Compression for the Mars Exploration Rovers) explain the need for data compression for the early 2004 Mars Exploration Rover (MER) mission. This led to the development of ICER, a state-of-the-art software wavelet-based image compressor. Image examples are provided as part of the explanation of the ICER features.

Todd A. Ely and Robert H. Bishop (Adaptive Interplanetary Navigation) are automating interplanetary navigation with a model adaptation solution. They provide an example using a filter bank, representing models of the Martian atmosphere for Mars entry, descent, and landing with a Hierarchical Mixture of Experts (HME) algorithm. This work appears promising for future missions.

Technology—Non-Deep Space
(a new category initiated with this issue)

Adrian Stoica (Evolvable Hardware) initiated evolvable hardware research at JPL in 1966. He describes spacecraft survival dependent on evolvable hardware for a spacecraft environment. An evolvable hardware example is an electronic circuit combined with a reconfigurable device. In the future, it is expected that spacecraft evolvable hardware will adapt to aging and radiation effects, allowing new functions for mission opportunities or changes.

Science

David Rochblatt and co-authors (Raster-Scan for Calibration of DSN Antennas) report on the application of a raster-scan technique to calibrate a 13.8-gigahertz receiving system that was built for the Cassini-Jupiter microwave observing campaign. The raster-scan subsystem performs on-the-fly mapping of extended and "pointlike" radio sources, produces three-dimensional maps of the source intensity, and reports a comprehensive set of antenna parameters that include aperture efficiency, pointing offset errors, focus errors, and the antenna beam shape. Preliminary results showing how these parameters vary as the DSS 13 beam waveguide antenna tracks in azimuth and elevation are reported.

**COMMUNICATIONS
SYSTEMS**

ANALYSIS
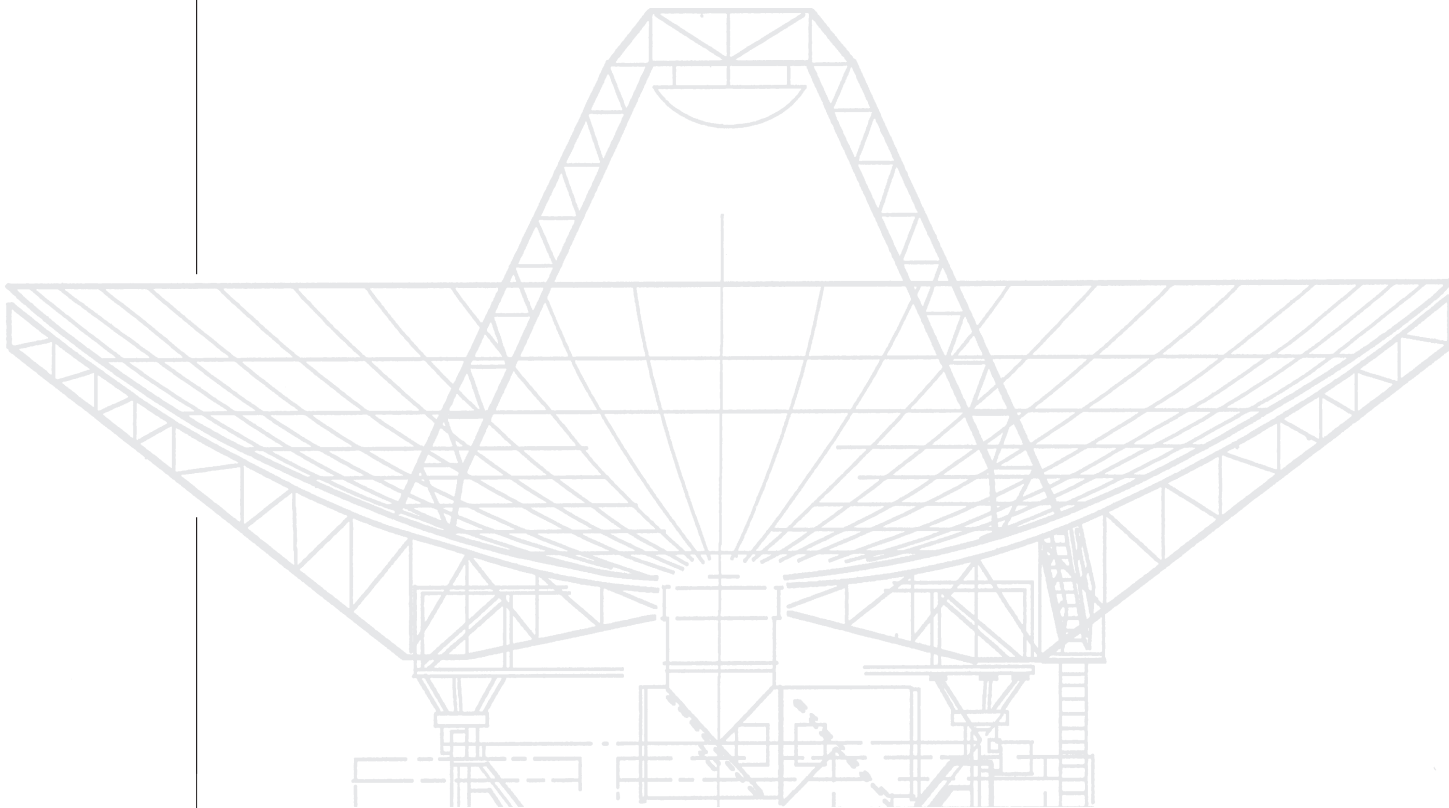
Deep Space Network Turbo Decoder Infusion
Enhanced Performance and Lower Decoder Complexity

Fabrizio Pollara and Kenneth Andrews

Introduction

Since their discovery in 1993, turbo codes have generated a great deal of interest due to their excellent performance and relatively low decoding complexity. In 1999 turbo codes have been included in the recommendations of the Consultative Committee for Space Data Systems (CCSDS) for telemetry channel coding and turbo encoders are being included in the flight hardware and software of missions in development. This article describes the effort to deploy turbo decoders in the Deep Space Network (DSN) to service missions launching in 2003 and later and the implications of these new capabilities for the design of future missions.

Turbo codes come very close to the attainment of the "Holy Grail" of information theory sought for half a century: the ultimate capacity of a communication channel as quantified by Claude Shannon in 1948. History shows that there is typically a substantial time gap between the formulation of a new error-correcting code and the implementation of decoders suitable for operational communication systems. The familiar convolutional codes were invented in the fifties; decoders for these codes became widely available in the seventies and were used for the Voyager mission. A similar timeline holds for Reed-Solomon codes invented in 1960 and used for Voyager's outer planets flybys. The gap became narrower for turbo codes invented in 1993, since our turbo decoder prototype was first tested in 2000. Commercial applications of turbo codes are also widespread, including their planned deployment in the upcoming third generation cellular phone system (Universal Mobile Telecommunications System).

Turbo Decoder Prototype

After a first attempt at designing a field programmable gate array (FPGA)–based turbo decoder to be used for a modified coding system emulating a turbo code on the Cassini spacecraft, it was decided to reduce the risk and cost of the first implementation and resort to digital signal processor (DSP) chips — the fastest on the market. Each DSP contains eight internal arithmetic units, so it can simultaneously perform as many as six adds and two multiplies. Collectively, the eight DSPs deliver an impressive 12,800 millions of instructions per second (MIPS).

The use of DSPs avoided the design of complicated and expensive hardware, as was the case with the previous generation of maximum-likelihood convolutional decoders (MCD III). The first goal was set at decoding the CCSDS turbo codes at 50 kilobits per second (kbps) per DSP chip. It took a lot of ingenuity and perseverance to attain this first goal, due to the intricate structure of the DSP. The DSP provides a certain degree of parallelism, but only for a selected subset of instructions. The fundamental routines for the computation of the forward and reverse passes on the trellis and the evaluation of the extrinsic reliabilities were all carefully hand optimized in assembly language, with painstaking attention to maintaining a high level of occupancy of each arithmetic unit, minimizing idle cycles, and carefully allocating memory. The first encouraging result was to realize a Viterbi decoder (which has a similar but simpler structure) with highly efficient use of the DSP. A lot more work led to the success in decoding a short turbo block. The next hurdle was to accommodate large block sizes without the need for large transfers of data between the DSP and external memory

because the memory available within the DSP was not sufficient to hold a whole decoding block. The concept of a "sliding window" or, more precisely of sub-blocks with some overlap "glue," proved to be essential for obtaining good decoding speed. The 50-kbps speed barrier was finally broken! This barrier has been more than doubled in the current decoder design.

The decoder prototype (see Figure 1) was implemented on two Versa Module Eurocard (VME) cards with four 200-megahertz DSP chips on each card. Six DSPs were used for the decoder itself, one for the synchronizer, and the last for peripheral services and for control (see Figure 2). The final decoder will use a single VME card with eight 300-MHz DSPs, achieving approximately 700 kbps (116 kbps per DSP) thanks to a modified algorithm using "stopping rules," as described later in this article.

A more detailed description of the turbo decoder can be found in [1] and [2].

Figure 2.
Turbo decoder
structure
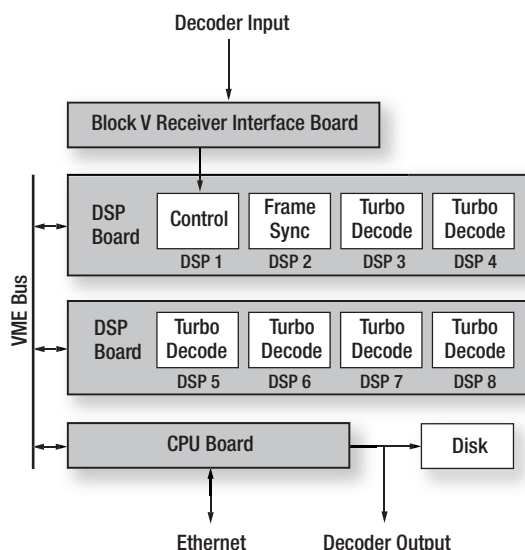
Technology Infusion and Interfaces to the DSN

An essential ingredient for the success of this task was the early involvement of designers familiar with DSN implementation issues. Throughout the development of the prototype and during the current implementation of the final units for the DSN there has been a close and continuous interaction between researchers, familiar with the turbo decoder principles, and implementers, familiar with interfaces and other DSN requirements. In this respect, the infusion of this technology has been exemplar. This interaction was also successful because of the ability of the different experts to speak each other's "language" and, in particular, to Jeff Berner's breadth of knowledge of DSN systems.

Besides the essential function of decoding the turbo-encoded data blocks and outputting the decoded bits in the CCSDS Standard Formatted Data Units (SFDU), the turbo decoder encompasses several peripheral but important functions: time-tagging, frame synchronization, pseudo de-randomization, and cyclic redundancy code (CRC) checking.

Time-tagging is accomplished by counting the cycles of the 10-MHz reference signal and appending an accurate time tag to each symbol. Next, frame synchronization is performed. The frame synchronizer searches for the frame-synchronization marker that was appended to the coded block. The frame synchronizer checks for both normal and inverted polarity in the marker; if the synchronizer detects the inverted polarity, the encoded block is marked for inversion before being sent to the decoding task. The synchronizer can buffer a minimum of four frames. This allows the system to acquire synchronization and then apply it backwards to the previous frames, reducing loss of data during the lockup period. Also, since the search for the marker is done in the symbol domain, the synchronizer operates at a low signal-to-noise ratio (SNR); therefore, it is necessary to utilize multiple frames to boost the SNR for sufficient synchronization reliability. A single DSP can do the job of matching four 192-bit sync patterns, giving optimum detection within the permitted latency. The synchro-

nized block is passed to the decoder element, along with a flag indicating whether or not the polarity of the block is inverted. Also, if pseudo randomization was applied to the code block, it is removed.

Performance and Decoding Speed

Turbo codes provide up to 0.9-dB improvement in E_b/N_o over the current best codes used by deep space missions and a substantial reduction in decoding complexity compared to that needed for the Galileo and Cassini codes. It is expected that the lower complexity and the modular design with multiple DSPs will greatly simplify the maintenance of the decoder units, in sharp contrast with the difficulties encountered in the troubleshooting of the ailing MCD III decoders. The performance of turbo codes compares favorably to that of current codes with equal codeblock size (see Figure 3). Note: E_b = energy per bit; N_o = single-sided noise power spectral density.

The initial decoder prototype used a fixed number of iterations. However, we have shown that simple schemes, called "stopping rules," can be used to decide to stop iterating based on some simple internal reliability measures. This decreases the average number of required iterations for a given data quality requirement and provides a useful tradeoff between power efficiency and decoding speed (see Figure 4).

The evolution of the speed performance of the turbo decoder (Figure 5) is compared to the speed of the current MCD III decoders for long-constraint-length convolutional codes used for Galileo, Cassini, Pathfinder, and other recent missions. Figure 5 shows the planned deployment of the turbo decoder for missions launching in 2003 and later. More recently, Mars missions such as the Mars Reconnaissance Orbiter (MRO) have an interest in higher data rates up to 4 or 5 Mbps. The trend seems to point to even higher data rate requirements in the near future. We are currently in the midst of negotiations aimed at realizing a next generation turbo decoder based on FPGA technology. This has been recommended as the best path for the evolution of DSN decoding services.
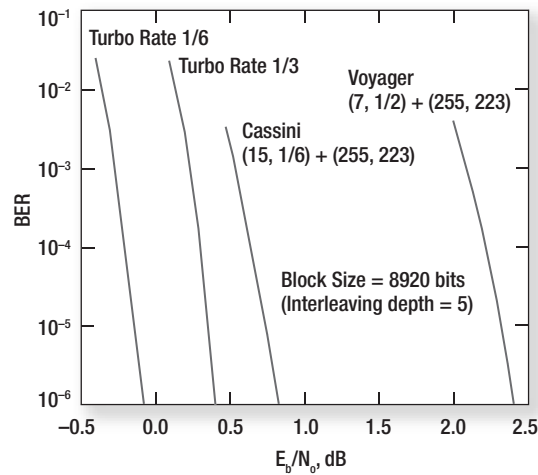

Figure 3. Performance of two turbo codes, compared with current codes of similar rate and block length
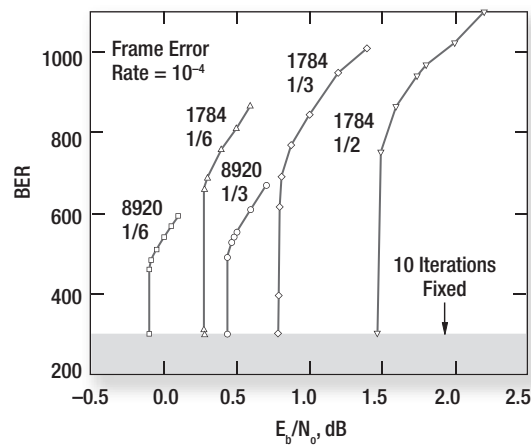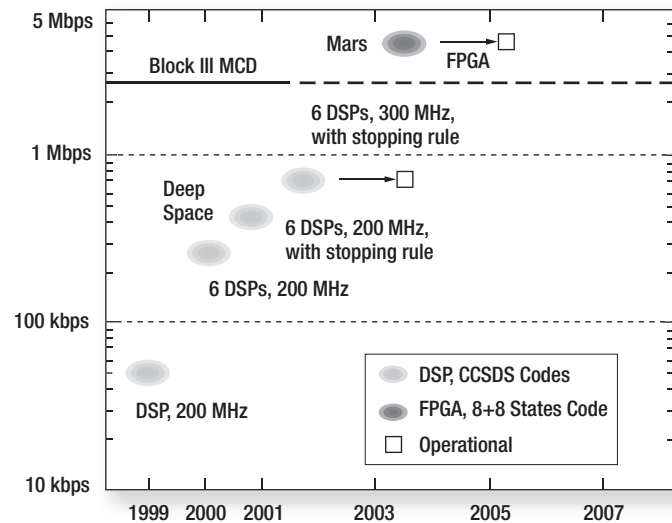

Figure 4. Tradeoff between power efficiency and decoder speed permitted by a stopping rule

This decision results from careful consideration of other alternatives, such as further increasing the number of DSP cards or relying on expected technological advances in speed to be provided by future DSP chips. If the FPGA turbo decoder project can be funded soon, we will be able to service MRO in 2006.

For decoding speeds in the tens of Mbps, we will probably have to resort to a different class of codes called low-density parity-check codes (LDPC), which are close relatives of turbo codes and as good as turbo codes in terms of power efficiency. LDPC have been predicted to be practically decodable up to 1 Gbps in recent studies.

When data rates are so high, it is also important to consider the spectral efficiency of the communications system (modulation and coding, combined). We are currently comparing the most promising bandwidth-efficient

5 Mbps

Block III MCD

Mars FPGA

6 DSPs, 300 MHz,
with stopping rule

1 Mbps

Deep
Space

6 DSPs, 200 MHz,
with stopping rule

6 DSPs, 200 MHz

100 kbps

DSP, CCSDS Codes
FPGA, 8+8 States Code
Operational
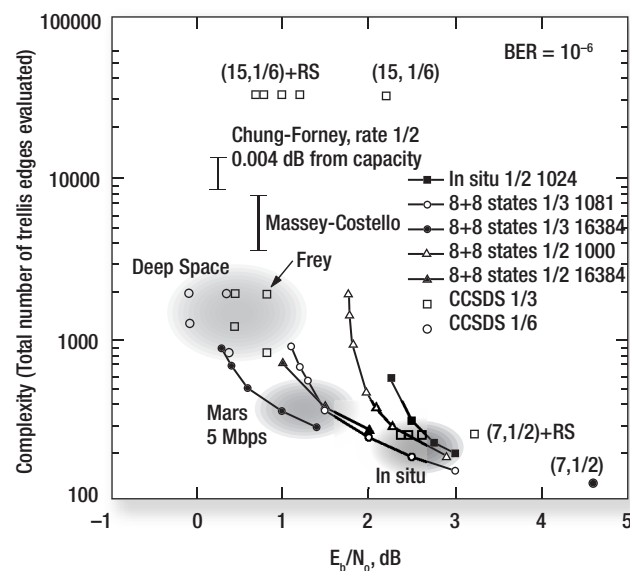
DSP, 200 MHz

10 kbps

1999 2000 2001 2003 2005 2007

coded-modulation schemes: coded Feher's quadraphase-shift keying (FQPSK) with iterative decoding similar to that used in serial turbo codes, or "pragmatic turbo codes" developed in FY01, using punctured turbo codes. The former scheme can deliver high spectral efficiency thanks to a special form of filtered QPSK signaling, but would require major modifications in the currently-planned DSN decoding system. The second scheme has the advantage of being decodable with the same turbo decoder described in this article, but has limitations due to its noncon-

stant-envelope signal set and related effects due to the power amplifier nonlinearity.

Throughout the development of new turbo coded systems for the DSN we have considered the best tradeoffs in terms of power efficiency of the code (required E_b/N_o for a given quality of the decoded bits or frames) and decoder complexity that translates into sustainable data rate. The tradeoffs in terms of these two main metrics are illustrated in Figure 6 for three scenarios: deep space mission (up to 700 kbps using the CCSDS turbo

100000

Complexity (Total number of trellis edges evaluated)

BER = 10^{-6}

(15,1/6)+RS (15, 1/6)

Chung-Forney, rate 1/2
0.004 dB from capacity

10000

Massey-Costello

In situ 1/2 1024
8+8 states 1/3 1081
8+8 states 1/3 16384
8+8 states 1/2 1000
8+8 states 1/2 16384
CCSDS 1/3
CCSDS 1/6

Deep Space Frey

1000

Mars
5 Mbps

(7,1/2)+RS

(7,1/2)

In situ

100

−1 0 1 2 3 4 5

E_b/N_o, dB

codes), Mars missions requiring high data rates (up to 5 Mbps), and *in situ* links. Deep space links can afford larger codeblock sizes and low code rates. Mars missions need to use higher code rates for spectral efficiency and may sacrifice a little performance for lower complexity.* *In situ* links require even lower complexity (the decoder is in the orbiter) and short codeblocks for bursts of communication. Onboard decoder development is currently not funded, but could substantially benefit from an FPGA-based decoder design since such a design can be easily ported to rad-hard FPGAs or commissioned to a very large scale integration (VLSI) foundry providing rad-hard processes.

Figure 6 also shows the substantially higher complexity incurred by the decoders for the Cassini/Pathfinder codes shown as (15,1/6)+RS. This figure includes some relevant results published by researchers elsewhere that we have painstakingly tracked and recorded.

Besides power and bandwidth efficiency, future codes will have to contend with ever increasing demands in terms of decoded data quality. Our typical design target of 10^{-6} bit error rate or 10^{-4} frame error rate may not be

sufficient for future links closely knit into a complex network. Improving the error-rate performance is difficult for turbo codes due to their characteristic behavior, called "error floor." However, we are studying methods for such improvements that will be based either on an additional outer code or on LDPC codes.

Epilog (Or a New Beginning?)

Will the discovery of turbo codes and the implementation of their decoders put an end to the saga of error-correcting codes? At first glance, this appears to be the case since their performance is so close to the ultimate limits; in fact, the reverse seems to be true. The "turbo principle" is having a profound influence on related disciplines and has led to something of a renaissance in communication theory research, including equalization, synchronization, and complex estimation problems in general.

References

[1] K. Andrews, V. Stanton, S. Dolinar, V. Chen, J. Berner, F. Pollara, "Turbo Decoder Implementation for the Deep Space Network." *IPN-ISP Progress Report* 42-148, Feb. 15, 2002.

[2] J. Berner, K. Andrews, "Deep Space Network Turbo Decoder Implementation," Aerospace Conference, Big Sky, Montana, June 2001.

Current negotiations for a future turbo decoder for Mars missions up to 5 Mbps seem to have reached a consensus to implement a faster decoder for the same CCSDS codes originally developed for deep space missions and to avoid the introduction of a new code for Mars.

A. Biswas,
M. J. Britcliffe,
D. J. Hoppe,
W. T. Roberts,
and
N. A. Page

**OPTICAL
COMMUNICATIONS**
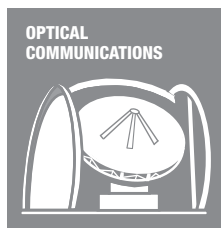
Large-Aperture Antenna Design for Deep Space Optical Communications

Introduction

NASA's Deep Space Mission Systems (DSMS) roadmap emphasizes enhanced telecommunications capacity, as one of the key thrusts for future technology development. Deep space optical communications is being pursued at JPL as a potential candidate technology that will provide the sought-after enhancement. With orders of magnitude bandwidth expansion, enhanced data transfer rates surpassing X- and Ka-band capability will be possible at optical frequencies. Furthermore, significant reduction of communications payload mass and volume can accompany the increased data rates. A ground-based receiving infrastructure, not unlike NASA's existing deep space network (DSN), but operating at optical frequencies, is required for an optical communications-based robust data service. A key element of the ground-based receiver is the large-aperture antenna (optical telescope) needed to collect and focus a sufficient number of signal photons transmitted from deep space. It is important to recognize that deploying optical receiving antennae in Earth orbit allows bypassing adverse weather and atmosphere-related effects on laser propagation. However, the cost for implementing orbiting receivers is prohibitive, coupled with the high risk for single point failure. Therefore, a critical milestone on the path to space-based optical receivers is a ground network that will cost-effectively validate, develop, and advance the technology. This intermediate solution will utilize a distribution of strategically geo-located optical antennae that will rely on weather diversity to overcome channel outages. Therefore, a cost-effective solution for large-aperture antennae is critical to the successful realization of a network capable of providing a high-quality optical-data service

from deep space. This report provides an update on recent conceptual designs that identify viable solutions for the large-aperture optical antenna.

Requirements

The large-aperture optical antenna should provide a sufficiently large collecting area and adequate optical surface quality so that robust optical communications links from interplanetary distances can be realized. Previous studies [1] have established that a 10-meter diameter aperture provides a good optimization between collection area and cost. Unlike conventional astronomical telescopes used for imaging, optical communications antennae do not need a constant wave-front error over a large field-of-view (FOV). The FOV should be large enough to accommodate the blind pointing accuracy, of the order of 100 microrads or less. However, because optical antennae must operate in both the day and nighttime, minimizing the FOV benefits the link performance by reducing background noise incident upon the detector. Therefore, strategies utilizing a larger, initial acquisition FOV that can be narrowed down once tracking lock with the spacecraft signal is achieved are favored. The actual spot size over which the collected signal energy is focused is the root sum square of the collective optical surface figure and the atmospheric "seeing." In the nighttime, under favorable atmospheric turbulence conditions, the overall blur size can be as low as 10 μrad, whereas daytime "seeing" will increase the blur size nominally 2 to 5 times. In addition to background noise contributions from within the FOV, highly efficient stray sunlight rejection must also be implemented in the design. Stray sunlight can either scatter into the focal plane, with severe diluting of signal strength, or contrib-

ute to undesirable thermal effects by incidence upon structural parts. For example, background light must not increase by more than 3 decibels (dB) as a result of pointing the antenna within 10 degrees of the Sun. Analysis indicates [2] that the bi-directional reflectance distribution function (BRDF) must be restricted < 0.003. This is not a particularly challenging requirement as BRDF values of 10^{-5} are routinely required for telescope parts.

Approach

In previous designs for the optical antenna, an aspheric primary and secondary mirror were considered. However, recognizing the FOV requirements for the optical antenna, a spherical primary mirror design was recently considered. The fabrication cost for a spherical compared to an aspherical primary mirror is a factor of 10 less. A 10-m spherical primary would comprise 100 identical 1-m-sized panels. An example of such a primary mirror is seen in the Hobby-Eberly Telescope (HET) at the Macdonald Observatory, run by the University of Texas [3], Austin, Texas. The surface finish and panel alignment requirements met by the HET design have been determined to exceed the requirements needed for an optical antenna. Considerable cost saving could also be realized in maintenance of a spherical primary: since all the panels are identical, they could be swapped using a low inventory of spare panels compared to individual unique panels for an aspheric primary mirror.

Optical design considerations, especially spherical aberration, were addressed in order to assess the suitability of a spherical primary mirror design. Two viable options were identified: a prime-focus and Cassegrain design. Figure 1 shows the layout for the prime focus design. Here, the secondary and tertiary mirrors are both aspherics; however, the additional costs of these smaller mirrors is expected to be more than offset by the cost savings expected from the use of a spherical primary mirror. Detailed optical modeling of this configuration shows that with an atmospheric "seeing" of 15 μrad, the blur size at the prime focus will be 110 ± 20 micrometers (μm).

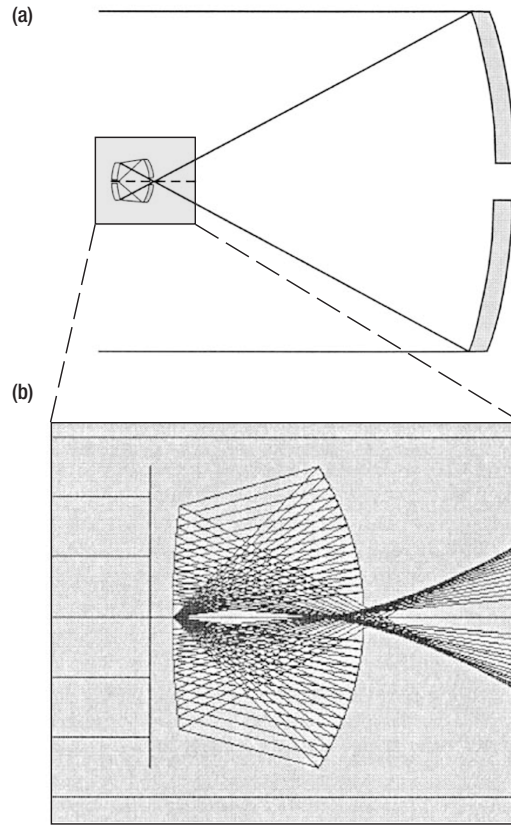The advantages of this design are that it is compatible with highly-efficient, stray-light

Figure 1.
(a) Prime focus configuration using a spherical primary mirror with a clam-shell corrector
(b) Expanded view of the corrector

rejection since the detector is pointed away from the stray-light sources outside the FOV. Furthermore, all light entering the secondary-tertiary assembly and ultimately reaching the detector must pass through the small, 110-millimeter (mm)–diameter hole in the tertiary mirror. The disadvantage of this design is the location of the detector assembly that does not provide easy access. The angle over which light is incident upon the detector is also high and, hence, conventional, narrow band-pass filters (NBPF) will suffer large transmission losses. This design will necessitate devising novel NBPF designs to overcome this restriction.

The Cassegrain design optical layout is depicted in Figure 2. Here, the secondary mirror is a high-order aspheric. This layout alleviates problems with detector and narrow band pass filter placement even though the surface figure of the secondary is more stressing from a fabrication standpoint.

Bulk glass, light-weighted glass, slumped glass, composites, and diamond-turned metallic surfaces were examined as candidate primary mirror panel materials. Based upon

optical surface quality, mechanical stiffness, primary mirror weight (influences support structure), and cost, light-weighted glass panels were selected [4]. Glass provides the best optical performance; in addition, the light-weighting lowers the complexity and cost of the mirror support structure. Furthermore, designing the system with minimum focal length reduces the overall optical path length. This simplifies the structural requirements and reduces the size of the dome. A 26-m geodesic dome was found to be adequate for the conceptual design. A conventional azimuth-elevation mount with a pointing and tracking accuracy of ± 50 μrad was identified for tracking of spacecraft at interplanetary ranges. Based upon the accuracies achieved in pointing the large DSN antennas, this requirement for a 10-m antenna should not be too stressful.

A number of additional considerations for the optical antenna design are sun shielding with good stray-light design of baffles and stops. Furthermore, the optical surfaces must be kept very clean since the deposition of contaminants on the surface will result in a degradation of the BRDF, causing what may prove to be an unacceptable increase in stray light. The dome will prevent exposure to airborne contaminants to a certain degree; however, a built-in, carbon dioxide (CO_2) snow mirror-cleaning system is highly recommended to alleviate mirror contamination problems. An active panel control is critical for maintaining blur circle size. Thermal effects and non-repeatable structural deformations will be the primary factors that cause misalignment of individual panels. The update rate of an active monitoring and correction system must have adequate bandwidth to deal with these effects. Analysis has shown that

with a charge-coupled device (CCD) camera located off-axis, a scheme based upon deliberately off-pointing individual panels will allow monitoring and correcting panel alignment.

Conclusion

The telescope design described addresses the key technical challenges of implementing a system for a future optical DSN. More rigorous optical tolerances and stray light analysis are needed to further refine and advance the strawman design summarized in this report. Figure 3 provides a view of the conceptual design discussed.
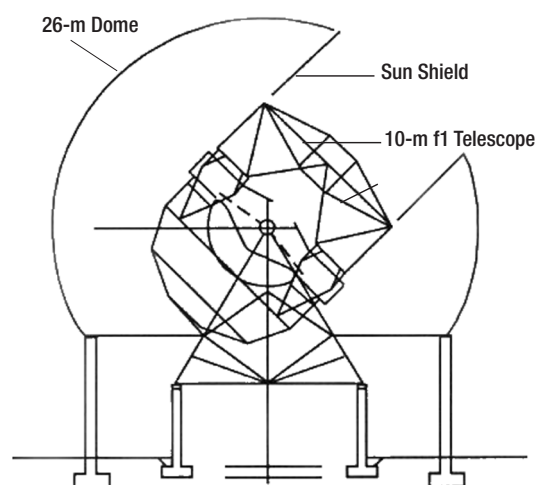
References

[1] J. R. Lesh and D. L. Robinson, "A Cost Performance Model for Ground-Based Optical Communications Receiving Telescopes," *TDA Progress Report* 42-87, July-September, 1986.

[2] M. Britcliffe, D. Hoppe, W. Roberts, N. Page, "A Ten-Meter Ground Station Telescope for Deep-Space Optical Communications: A Preliminary Design," *IPN Progress Report* 42-147, Nov. 15, 2001.

[3] http://www.as.utexas.edu/mcdonald/het/het.html

[4] M. J. Britcliffe and D. J. Hoppe, "Main-Reflector Manufacturing Technology for the Deep Space Optical Communications Ground Station," *TMO Progress Report* 42-145, May 15, 2001.

Ka-Band Traveling Wave Tube Amplifier

Inam Haque
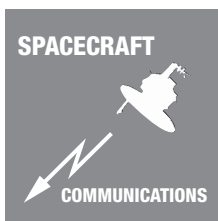
Summary

The increased number of space missions and the increased sophistication of the mission instruments means we need to transmit increased engineering and scientific data back to Earth at a faster rate. One solution is to utilize Ka-band (around 32 gigahertz [GHz]), which allows a two-to-six-fold increase in data rate over X-band (around 8.4 GHz). Feasibility of the Ka-based downlink has been demonstrated in current space flight experiments.

JPL has recently managed the development of a 35-watt (W), Ka-band Traveling Wave Tube Amplifier (TWTA) using up-to-date advances in proven technology. This high efficiency, long-life TWTA is designed to be compatible with current and future spacecraft and to live through long radiation exposure. It will operate through extreme temperatures and severe spacecraft vibration without degradation in performance.

Background

The number of spacecraft and probes engaged in the scientific exploration of space is on the rise, especially for missions to Mars, and the need to transmit engineering and scientific data back to Earth at a faster rate is now at a critical threshold of current equipment capability. Advancing instrument sophistication and capability has increased the sheer quantity of science data collected in missions. One solution for increased scientific data return is to utilize Ka-band as the primary means to communicate/transmit data. Ka-band provides a realizable two-to-six-fold increase in data rate over currently used X-band. Feasibility of the Ka-based downlink has been demonstrated by the Mars Global Surveyor Ka-band Link Experiment (KaBLE II) and by the Cassini Ka-band Radio Frequency Instrument System operation.

To provide an incentive for program offices and planned missions to switch to Ka-band, the Interplanetary Network Directorate (IND) Technology Office has been investing in the development of a 35-W Ka-band TWTA since 1998. The TWTA (see Figure 1) consists of a TWT plus a high-voltage power supply (HVPS). An engineering qualification model and a proto-flight model of the TWTAs are due for delivery to JPL by mid-2002. The actual development of the TWTA has been subcontracted to Thales Electronique (formerly Thomson Electronique) who has partnered with Tesat Spacecom (formerly Bosch SatCom) for the HVPS.

As a result of this IND investment, the 2005 Mars Reconnaissance Orbiter (MRO) mission will be using this amplifier in a backup communications system to the primary X-band system. Analysis shows that a Ka-band telecom system (radio, TWTA power amplifier, microwave components, and antenna) with a 35-W, Ka-band TWTA would perform about the same task as an X-band telecom system with a 100-W, X-band TWTA. It is anticipated that future missions will adopt Ka-band as the primary mode of communication and transmittal of scientific and engineering data.

Design and Requirements

The 35-W, Ka-band TWTA uses new advances in proven technology. The key parameter, power output, is controlled by a self-adjusting anode control voltage, A_0, that maintains a constant beam current. This TWTA is designed to be adaptable to any 28-volts direct current (Vdc) platform and

Figure 1.
Ka-Band
TWTA block
and
interface
diagram
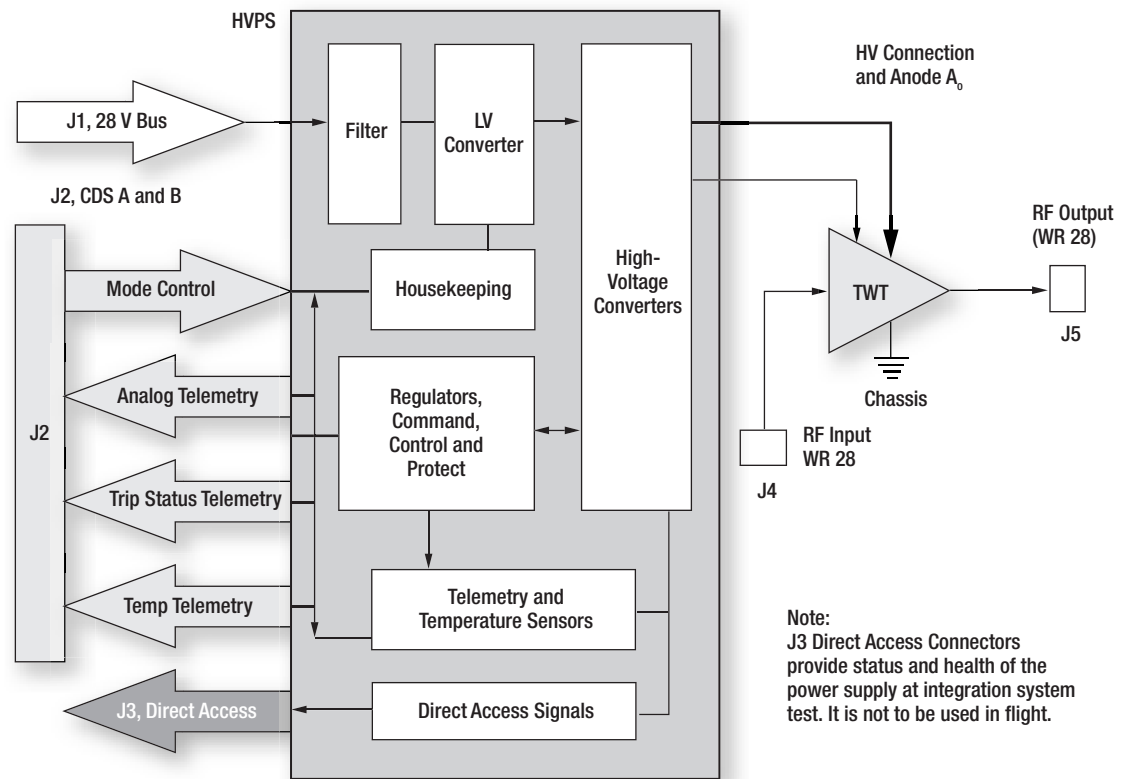
compatible with currently used spacecraft telemetry, electromagnetic compatibility (EMC), and grounding (see Figure 2). It has a very-high efficiency (engineering model measurements indicate greater than 46%), is designed for long mission life (12 years or more), and will live through 200 kilorads (krad) total dose radiation. An added fea-

Figure 2.
Ground
scheme
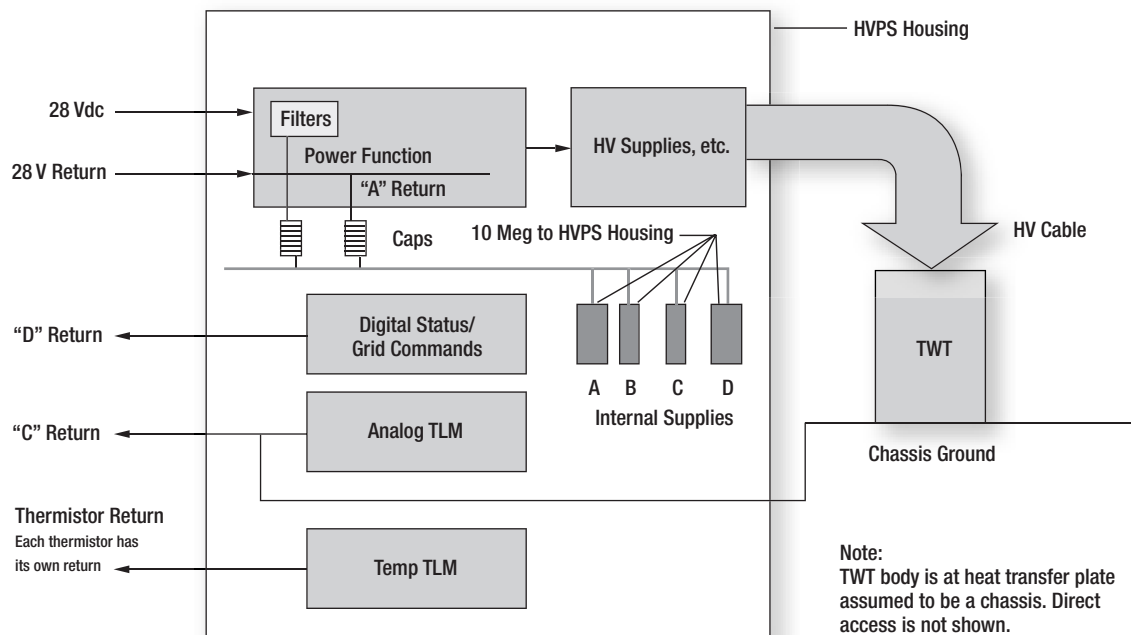
ture of this TWTA is that it will operate through temperature extremes of −40 degrees Celsius (° C) to +75° C and will operate through random vibration up to 19.6 root mean square acceleration (Grms) without degradation in performance. This robust design assures that the TWTA performance will not degrade during normal launch, cruise, orbital, or landing maneuvers.

Looking at the history and evolution of TWTA development, it has been a struggle to unify TWTA requirements that meet all possible scenarios of performance over environmental extremes (temperature, radiation, vibration, pressure) while trying to meet requirements during launch, course correction, orbital maneuvers, or entry/descent/landing where downlink is needed. The importance of operation during maneuvers was evident during the investigation of the lost Mars Climate Orbiter and the Mars Polar Lander spacecraft. The current operational TWTAs in space are not capable of operation during spacecraft vibration created by maneuvers or course correction. Mission Control goes through anxious moments when maneuvers are carried out with the TWTA powered off and no downlink. There is a consensus that a power output of 35-W Ka-band and 100-W X-band are considered as the appropriate level of output power for near-term future Mars orbiter missions.

JPL developed a new set of generic requirements for this development that will simplify the spacecraft interface to the TWTA and maximize capability and performance for future missions. The requirements are specified for a 28-Vdc platform as follows, but are equally applicable or adaptable to higher voltage platforms.

- End-of-life radio frequency (RF) power output shall be 35 W.

- The TWTA power consumed at saturation shall be no greater than 77 W (goal = 74).

- The TWTA shall have efficiency greater then 46% to minimize power dissipation when in saturation.

- The TWTA shall operate in a linear mode with minimal phase shift and third-order modulation when operated with two carriers.

- The TWTA mass shall not exceed 2.77 kilograms (kg), including the high-voltage cable.

- The TWTA RF drive shall be sufficiently low to permit the use of "off-the-shelf" RF sources.

- The TWTA shall operate successfully without shutdown during normal cruise, orbital, or landing maneuvers with vibration up to 19.6 Grms. (The TWTA is off during launch.)

- The TWTA shall operate through "critical" pressure during Mars low-orbit maneuvers, Mars entry/descent/landing, or during Mars-landed operations.

- The TWTA shall provide all needed analog telemetry to determine the status of the TWTA critical functions, including body current, bus voltage input current, under voltage, and high body current.

- The TWTA shall operate to full performance levels between −40° C to +75° C where the spacecraft will see an extreme environment.

- The spacecraft grounding shall be simplified. The TWTAs control/command interface and status signal shall be isolated from each other and from the power bus to minimize ground interaction. This concept is shown in Figures 1 and 2.

- The TWTA shall be fully compliant to EMC requirements of Military Standard 461 (MIL-STD-461).

Development Status

Both a 27-W and a 35-W breadboard TWTA have been developed and tested by Thales (see Figure 3). The environmental quality model (EQM) is currently in build and test and is scheduled to be delivered to

Figure 3.
Outline
of the
TWTA
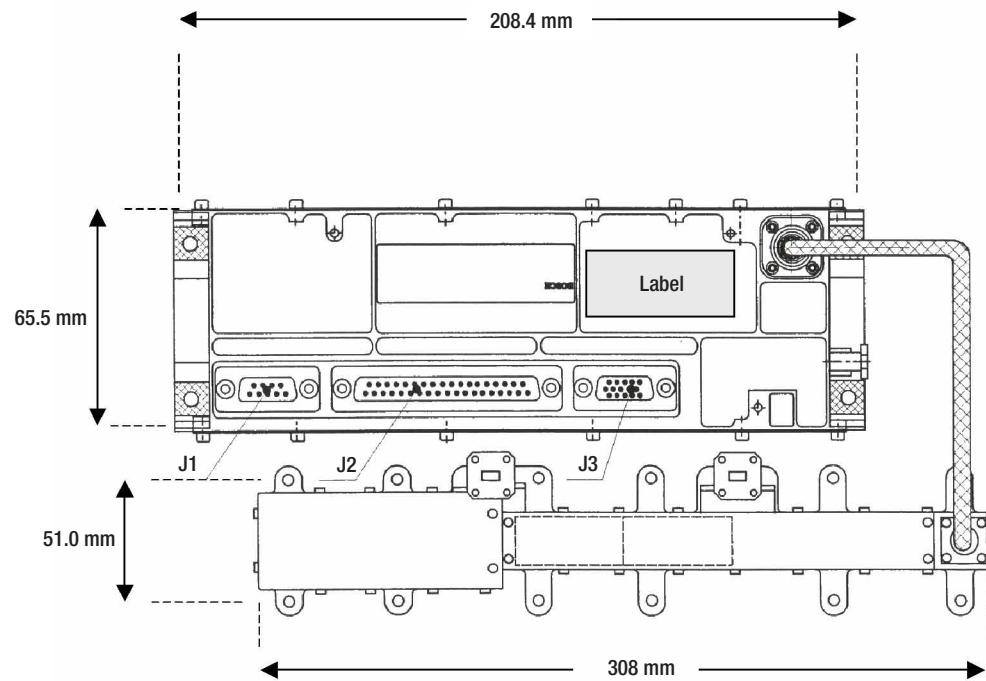
JPL in July, 2002. The preliminary design re-
view was conducted in April, 2001; a com-
bined IND and MRO critical design review
was completed in March, 2002. All the re-
quired analyses have been completed and are

awaiting JPL approval, expected by June, 2002.
    Performance data on the breadboard TWT
is shown below and is better then expected.
The overall efficiency is slightly lower then
expected, but within reasonable limits.

| Frequency (GHz)	Input Power (dBm)[1]	Output Power (dBm)	Output Power (W)	Power Dissipated (W)	Power Consumed[2] (W)	TWT Efficiency %
31.75	−5.0	45.60	36.3	30.6	69.64	52.14
32.00	−4.1	45.66	36.8	30.4	69.74	52.76
32.25	−4.4	45.61	36.4	30.2	69.24	52.57

[1] dBm: power levels above or below 1 milliwatt.
[2] The delta power between consumed and dissipated RF output is the filament power of about 2.71 W.

Conclusion

    Promising development performance
achieved to date and selection by MRO to fly
the 35-W Ka-band TWTA are major steps
by NASA and JPL to establish Ka-band
communications in space. Future missions to
Mars and other exploration flights may now
consider using Ka-band communication for
enhanced data transmission and collection.
As part of this emphasis, the NASA Glenn

Research Center has funded the Hughes
Electron Dynamics division of Boeing to
begin development of the next generation
100-W Ka-band TWT. JPL is planning to
develop an HVPS based on its successful
CloudSat HVPS that could power the
100-W Ka-band TWT when it is ready. As
missions to Mars, other planets, asteroids,
and comets increase, Ka-band will increase
NASA's communication capability.
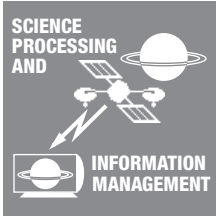
ICER on Mars
Wavelet-Based Image Compression for the Mars Exploration Rovers

Aaron Kiely, Matthew Klimesh, and Justin Maki

Introduction

In early 2004, the Mars Exploration Rover (MER) mission will land a pair of rovers on Mars and operate each of them for 90 Martian days. Each rover is equipped with an unprecedented nine visible-wavelength cameras: a mast-mounted, high-angular-resolution color imager for science investigations (the panoramic camera, or Pancam); a mast-mounted, medium-angular-resolution camera for navigation purposes (the Navcam); a set of body-mounted, front-and-rear cameras for navigation hazard avoidance (the Hazcams); and a high-spatial-resolution camera (the Microscopic Imager) mounted on the end of a robotic arm for science investigations. With the exception of the Microscopic Imager, all of these cameras are actually stereo camera pairs.

Not surprisingly, collecting and transmitting images to Earth will be a major priority of the mission: well over half of the bits transmitted from the rovers will consist of compressed image data from these cameras. For image compression, MER will rely exclusively on the ICER image compressor, developed in the Science Processing and Information Management work area at the Jet Propulsion Laboratory (JPL).

The development of ICER was driven by the desire to achieve state-of-the-art compression performance with software that meets the specialized needs of deep space applications. ICER's compression effectiveness will enhance the ability of the MER mission to meet its science objectives. In this article, we give a brief overview of the features and inner workings of ICER.

ICER Overview

ICER is a *wavelet-based* image compressor that features *progressive compression*. The ICER software can provide both lossless and lossy compression and incorporates a sophisticated error-containment scheme to limit the effects of data loss.

Progressive compression means that as more compressed data are received, successively higher quality reconstructed images can be reproduced, as illustrated in Figure 1. Thus, one could, for example, send a small fraction of the data from an ICER-compressed image to get a low-quality preview and later send more of the data to get a higher quality version if the image is deemed interesting. In future missions, progressive compression will enable sophisticated data return strategies involving incremental image quality improvements to maximize the returned science value using an on-board buffer.

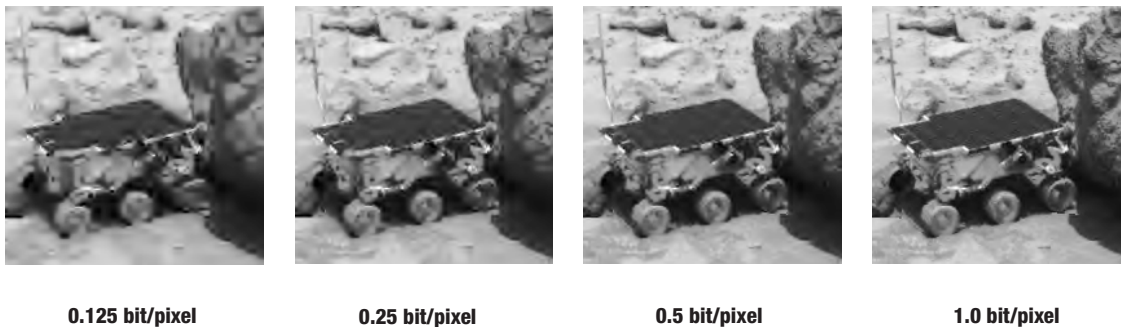| 0.125 bit/pixel | 0.25 bit/pixel | 0.5 bit/pixel | 1.0 bit/pixel |

Figure 1. This sequence of image details from a larger image shows how image quality improves under progressive compression as more compressed bits are received

A wavelet transform is a linear (or nearly linear) transform designed to decorrelate images by local separation of spatial frequencies. In wavelet-based image compression, a wavelet transform is applied to the image before further processing. By using a wavelet transform, ICER avoids the "blocking" artifacts that can occur when the discrete cosine transform (DCT) is used for decorrelation, as in the Joint Photographic Expert Group (JPEG) compressor used on the Mars Pathfinder mission. Also, wavelet-based compression is usually superior to DCT-based compression in terms of quantitative measures of reconstructed image quality.

When used for lossless image compression, ICER produces compressed image sizes that are competitive with those produced by state-of-the-art algorithms designed exclusively for lossless image compression (e.g., [7]) and about 20% smaller than the Rice compressor used for lossless image compression on Mars Pathfinder.

In a nutshell, here's how ICER works. A wavelet transform decomposes the image into several subbands, each a smaller version of the image, but filtered to contain a limited range of spatial frequencies. ICER achieves progressive compression by successively encoding bit planes within the subbands. (A bit plane consists of all of the bits of a given significance from the subband; for example, the second bit plane of a subband contains the second most significant bit from each member of the subband.) While encoding, ICER maintains a statistical model of the image that allows effective compression of the bit planes. Lossless compression is achieved when all of the compressed subband bit planes are transmitted.

For error containment purposes, the compressed data are organized into a user-selectable number of segments, each containing information about a limited rectangular region of the image. Each segment is compressed independently so that if one of the packets for a segment is lost due to channel noise, the other segments are unaffected.

Controlling Image Quality and Amount of Compression

All MER cameras produce 1024-pixel by 1024-pixel images at 12 bits per pixel. Images transmitted from MER will range from tiny 64×64 "thumbnail" images up to full-size images. MER will make use of compressed bit rates ranging from less than a bit/pixel up to rates yielding lossless compression. Current MER baseline plans call for navigation, thumbnail, and many other image types to be compressed to approximately 1 bit/pixel; lower bit rates may be used for certain wavelengths of multi-color panoramic images. At the other end of the compression spectrum, radiometric calibration targets are likely to be compressed to about 4 to 6 bits/pixel, and lossless compression will be used for certain science images.

To control the image quality and amount of compression in ICER, the user specifies a *byte quota* (the nominal number of bytes to be used to store the compressed image) and a *quality level* parameter (which is essentially a quality goal). ICER attempts to produce a compressed image that meets the quality level using as few compressed bytes as possible. ICER stops producing compressed bytes once the quality level or byte quota is met, whichever comes first.

This arrangement provides added flexibility compared to compressors (like the JPEG compressor used on Mars Pathfinder) that provide only a single parameter to control image quality. Using ICER, when the primary concern is the bandwidth available to transmit the compressed image, one can set the quality goal to lossless and the given byte quota will determine the amount of compression obtained. At the other extreme, when the only important consideration is a minimum acceptable image quality, one can provide a sufficiently large byte quota and the amount of compression will be determined by the quality level specified.

Wavelet Transform

The first step in ICER compression is to perform a two-dimensional wavelet transform of the image. This transform decomposes the image into a user-controlled number of sub-

bands. To conserve memory on board MER, the wavelet-transformed image is stored using the same memory array as the original image.

The wavelet transform is used to decorrelate the image, concentrating most of the important information into a small number of small subbands. Thus, a good approximation to the original image can be obtained from a small amount of data. In addition, the subbands containing little information tend to compress easily. The wavelet transform does leave some correlation in the subbands, so compression of the subband bit planes uses predictive compression to attempt to exploit as much of this remaining correlation as possible.

The two-dimensional wavelet transform is accomplished through several applications of a one-dimensional wavelet transform to rows and columns of data. An ICER user can select one of several reversible integer wavelet transforms [1, 5]. Each is a nonlinear approximation to a linear high-pass/low-pass filter pair. The wavelet transform produces integer outputs and is exactly invertible, which allows us to achieve lossless compression when all of the subband data are reproduced exactly.

After the wavelet transform, ICER successively compresses bit planes of the subbands. Ideally, at each stage of compression, we would like to encode the subband bit-plane giving the biggest improvement in some measure of image quality per compressed bit. However, performing this optimization on-board would be impractical and would achieve rather modest gains in compression effectiveness; therefore, in practice, ICER uses heuristics to select which subband bit plane to transmit at each stage.

Context Modeling

ICER employs a technique known as *context modeling* in its encoding of the bit planes. With this technique, before encoding a bit in the transformed image, the bit is classified into one of several *contexts* based on the values of previously encoded bits. These bits are from the pixel being encoded and nearby pixels and include the bits previously encoded from the current bit plane as well as bits from previous (more significant) bit planes. The probability that the bit to be encoded is a "0" is estimated based on the encoder's previous experience with bits classified into the same context. The bit is then encoded based on this probability estimate. Since the probability estimate relies only on previously encoded information, the decoder can duplicate this calculation and produce the same probability estimate, which is essential for proper decoding.

Although the above procedure may sound complicated, in fact it is done with reasonably low complexity. In particular, ICER's scheme for classification into contexts is fairly simple: it uses two small lookup tables, is based on simple properties of the pixel being encoded as well as the 8 nearest neighbors pixels, and yields one of 18 contexts. This scheme is similar to the context modeling used by JPEG2000 [6].

Entropy Coding

The actual compression of the bits is accomplished with an *entropy coder*. An entropy coder is a module that takes a sequence of bits along with corresponding probability-of-zero estimates for the bits, and produces an encoded (and hopefully compressed) bit stream from which the original sequence of bits can be recovered. At the decoding end, it's necessary for the decompressor to reconstruct the probability-of-zero estimate for each bit before decoding it.

Although binary entropy coding is usually performed with a technique known as *arithmetic coding* (or with a low-complexity approximation to arithmetic coding), we have chosen to use a lesser known technique called *interleaved entropy coding* [2, 3, 4]. We have produced a software implementation of interleaved entropy coding that has particularly low complexity; therefore, it is well-suited for space applications and other applications where speed can be of critical importance. Given perfect probability-of-zero estimates, both arithmetic coding and interleaved entropy coding can compress stochastic bit sequences to within 1% of optimal.

Although the entropy coding stage of ICER is extremely effective, improvements to

the overall compression effectiveness should still be possible. The wavelet transform and the context modeling are designed to exploit as much correlation as possible from the image; however, this process is not an exact science.

Error Containment

Error containment is essential to accommodate bit error rates seen on the deep space channel. Without error containment, a single packet loss due to channel errors can corrupt large segments of compressed data.

To achieve error containment, ICER produces the compressed bitstream in separate pieces or *segments* that can be decoded independently. These segments represent rectangular regions of the original image, but are defined in the transform domain. If the image were partitioned directly and the wavelet transform separately applied to each segment, under lossy compression the boundaries between segments would tend to be noticeable in the reconstructed image even when no compressed data is lost, as illustrated in Figure 2(a). By segmenting the image in the transform domain, we can virtually guarantee that such artifacts will not occur, as illustrated in Figure 2(b). There are also secondary benefits: we achieve better decorrelation by applying the wavelet transform to the entire image at once and it is easier to maintain a similar image quality in the different segments. A minor side effect is that the effect of data loss in one segment can appear to "bleed" slightly into adjacent segments. Note that segments that are easier to compress will naturally use fewer bits, so lengths of compressed segments will be variable.

ICER includes an algorithm to automatically partition the image into the number of segments specified by the user. The segments

produced have two desirable properties. First, the segments tend to be nearly square. A square region has a smaller perimeter than an elongated region of the same area, so its pixels have more neighbors with which correlations can be exploited and, thus, more effective compression is possible. Second, the segments tend to have nearly equal areas, which makes it somewhat more likely that compressed segments will have similar lengths; therefore, data losses are less likely to corrupt larger regions of the image. Figure 3 illustrates the partitioning of a square image into 8 segments.

Dividing an image into a large number of segments can confine the effects of a packet loss to a small area of the image; however, it's generally harder to effectively compress smaller image segments. Since ICER provides flexibility in choosing the number of segments, compression effectiveness can be traded against packet loss protection, thereby accommodating different channel error rates. Note also that more segments are not always bad for compression effectiveness: many images are most effectively compressed using 4 to 6 segments because disparate regions of the image end up in different segments.

Because ICER is progressive, some error containment automatically occurs within segments as well: when a packet loss occurs, any previously received packets for the affected segment will allow a lower fidelity reconstruction of that segment, as illustrated in Figure 3.

Conclusion

With an armada of 18 cameras set to land on Mars in 2004, MER will rely heavily on ICER to deliver data back to Earth during the 180 Martian days of surface operations. Since ICER is a fairly general purpose algo-

Figure 2. Image details that illustrate (a) artifacts at segment borders that would arise if the wavelet transform were separately applied to each segment, and (b) elimination of such artifacts in ICER-compressed image


(a)

(b)
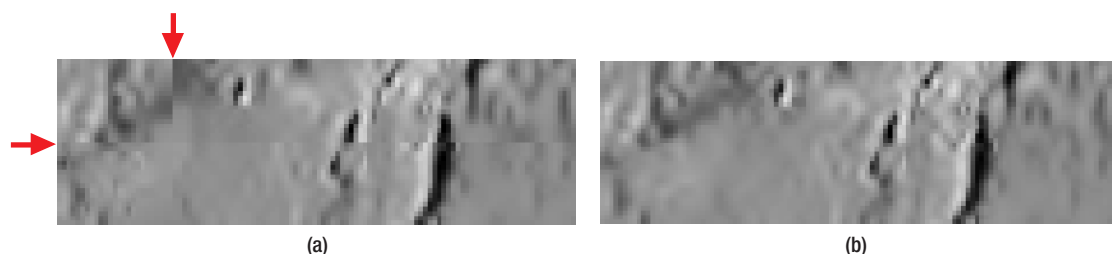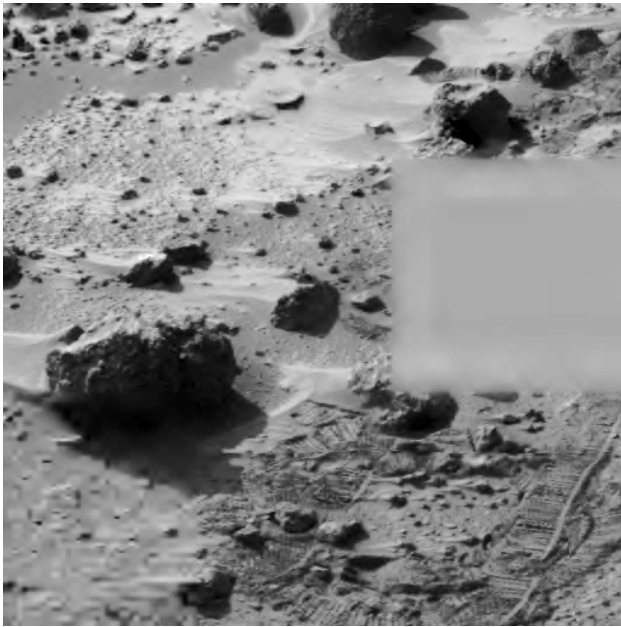
Figure 3.
Example of error containment in an image compressed to 1 bit/pixel, and suffering packet losses affecting three of eight image segments

rithm and provides good lossy and lossless image compression combined with effective error containment, it has the potential to benefit other missions as well. Rapid advancements in imaging technology will continue to push the need for innovative data compression technologies such as ICER. As the techniques improve and hardware speeds increase, future missions will be able reap larger benefits from image compression.

References

[1] M. D. Adams and F. Kossentini, "Reversible Integer-to-Integer Wavelet Transforms for Image Compression: Performance Evaluation and Analysis," *IEEE Transactions on Image Processing*, Vol. 9, No. 7, pp. 1010–1024, June, 2000.

[2] P. G. Howard, "Interleaving Entropy Codes," *Proc. Compression and Complexity of Sequences* 1997, Salerno, Italy, pp. 45–55, 1998.

[3] A. Kiely and M. Klimesh, "A New Entropy Coding Technique for Data Compression," *IPN Progress Report* 42-146, pp. 1–48, April-June, 2001.

[4] A. Kiely and M. Klimesh, "Memory-Efficient Recursive Interleaved Entropy Coding," *IPN Progress Report* 42-146, pp. 1–14, April–June, 2001.

[5] A. Said and W. A. Pearlman, "An Image Multiresolution Representation for Lossless and Lossy Compression," *IEEE Transactions on Image Processing*, Vol. 9, No. 5, pp. 1303–1310, September, 1996.

[6] D. Taubman, "High Performance Scalable Image Compression with EBCOT," *IEEE Transactions on Image Processing*, Vol. 9, No. 7, pp. 1158–1170, July, 2000.

[7] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I Lossless Image Compression Algorithm: Principles and Standardization into JPEG-LS," *IEEE Transactions on Image Processing*, Vol. 9, No. 8, pp. 1309–1324, August 2000.
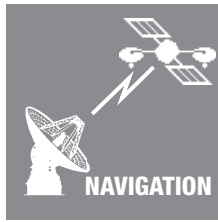
Adaptive Interplanetary Navigation

Todd A. Ely
and
Robert H. Bishop

Introduction

Spacecraft navigation possesses a dual nature: it is simultaneously a science and an art. The reasons for this are manifold; however, a common denominator to each is that the knowledge of the models that govern a spacecraft's motion and the instruments that detect this motion are imperfect. The navigator faced with this attempts to reconstruct and predict the motion (via trajectory determination), all the while knowing that the models at his/her disposal are not always adequate to the task. In order to improve their lot, navigators continuously seek to improve the models that they use by adapting them to the changing circumstances that the spacecraft encounters. Of course, traditional navigation algorithms (a.k.a., filters) are capable of adjusting internal model parameters using spacecraft measurement data (i.e., Doppler tracking, accelerometer readings); however, in certain situations, these models cannot be determined well. For instance, a spacecraft that maintains its solar arrays pointed towards the Sun experiences a solar radiation pressure that can appear as thrust in that same direction. If the filter simultaneously mismodels the solar array area and estimates thrust accelerations, it will inappropriately assign the acceleration to the thrust. Now, if the spacecraft subsequently points away from the Sun, any prediction of a future trajectory will be in error. Of course, a navigator will perform numerous fits, examine sensitivities, and, most likely, determine a filter model that better matches the observed measurement data. However, this process is iterative, takes time, and, ultimately, relies on the expertise of the navigator to isolate the mismodeling. It is a natural question to ask if the process of searching a design space can be automated.

Are there techniques to aid the navigator in this search? These questions lead us to investigate model adaptation solutions that yield answers in a post-processing environment (i.e., trajectory reconstruction and prediction).

Another equally important application of adaptive navigation is model detection, isolation, and resolution in a real-time or near-real-time autonomous mode. These capabilities reduce risk and increase robustness of navigation in environments that are uncertain and highly dynamic. An autonomous navigation capability can detect environment and spacecraft changes automatically and adjust filter models to accommodate the change and/or provide timely alerts to a ground system operations team of the need to make changes. Cruise, approach, aerobraking, or on-orbit operations are manually intensive mission phases that often require around-the-clock monitoring of spacecraft navigation; an autonomous monitoring (potentially, correcting) component eases the burden on operational teams. There is also a category of mission-critical events that occur over a very short time and, most likely, without the possibility of intervention from the ground. An example of this type of mission phase is atmospheric entry, descent, and landing. At Mars, this entire mission phase typically takes less time than is needed for a signal to return to Earth. Clearly a robust navigation system is required for mission success. In this case, an adaptive navigation algorithm could provide trajectory estimates directly to a guidance system for retargeting.

Adaptation in these two arenas, post-processing and real-time processing, lead to separate adaptive navigation approaches. However, all the techniques that we have in-

vestigated rely heavily on results from soft computing, using methods such as genetic algorithms and neural networks. They are nonlinear, very robust, simple to implement, and can accommodate varied sources of information. The Interplanetary Network Directorate (IND) has been funding our work in adaptive navigation to examine the feasibility of these techniques and build prototype tools to examine their performance on realistic problems.

Adaptive Navigation Using Genetic Algorithms for Post-Processing Applications

The problem of tuning trajectory determination models requires an intensive search of multiple dynamical and nondynamical models that yield trajectory solutions with minimal errors. Often, the tools at the disposal of the navigator have modeling limitations. As an example, the Mars Pathfinder (MPF) spacecraft had a backshell shroud that protected the lander during cruise and Mars entry. It was conical in shape and, during the cruise phase, was often shadowed from the Sun by the spacecraft solar arrays. However, the component models available for use by the MPF analysts for solar radiation pressure (SRP) modeling consisted of only flat plates and cylinders. Neither was entirely correct; thus, analysts were forced to approximate the model of the backshell using one of these choices with an unknown scale factor (because of shadowing). Their approach was iterative, requiring the team to select a model, adjust the filter realization, process the observations, and then compare results to previous filter realizations. Clearly, this tuning process could have benefited from a systematic and automated methodology for finding a best filter model. Doing so would have eased operation team workloads and allowed them to consider a wider range of possible solutions with a resultant better accuracy. As an example, the MPF team considered five different SRP models. The adaptive technique, to be described, examined nearly 4000 SRP models.

The experience of the MPF navigation team at finding an adequate SRP model illustrates the complexity of the navigation model design space and the ad hoc methods currently utilized to obtain adequate models. It should be emphasized that the nature of the team's design decisions focused not only on filter parameter selection and their associated values, but also on selecting the underlying dynamic models (a structural change to the filter). Past approaches to adaptive filtering have typically focused on filter tuning by considering only modifications to model parameters values and/or a priori uncertainties. Structural changes to filters via modifications to underlying dynamic models, measurement models, and/or filter-state vector components have been outside of the scope of most adaptive techniques. This is partially attributable to the fact that many design optimization methods are gradient based and, thus, do not support discrete structural changes. A genetic algorithm (GA) does not require gradient information and is not limited by a small sample of the desired design space; hence, it is a natural choice as a model optimization method.

A GA is a computational representation of natural selection that bases its search and optimization on the analogy that an individual that is more fit to its environment is closer to an optimal design [1]. In applying this analogy to the adaptive navigation problem, an individual represents a specific filter realization of dynamic models, measurement models, state vector components, a priori uncertainty values, and associated noise process models. The particular variables and their associated range of values that have been isolated for analysis represent the design space. Each individual design has its variable values encoded into a representation — typically a binary string — that corresponds to its chromosome. The GA examines a population of individuals by analyzing the fitness of each individual, a metric that measures a selected figure of merit. In the case of adaptive navigation, the fitness value is based on the sample statistics of the navigation filter's residual sequence. (Residuals measure the difference between the real and modeled tracking data.) With this information, the GA iterates on the members of the population from one generation to the next with the aim of im-

proving the overall fitness of the population, where improvement is defined in this problem as minimizing the fitness value.* The GA accomplishes this objective using three primary operators:

1. Selecting the more fit individuals of the population to become parents.

2. Mating the parents via a crossover operation that exchanges chromosome information to produce children.

3. Modifying the children's chromosomes via a low probability of mutation to ensure diversity of the population.

The children of the current generation then become the population for the next generation and the process iterates until convergence or until it is stopped. Furthermore, to ensure that the GA starts with a diverse image of the design space, the initial population is seeded with a random set of individuals. This combination of deterministic rules for selection and crossover coupled with probabilistic sampling creates a robust approach for arriving at the global minimum.

A general method for analyzing the fidelity of a filter that is processing measurement data is to examine the quality of its residual sequence. A properly operating filter should produce a residual sequence that is a zero-mean, white-noise process [2]. A sequence of residual values that results from operating a filter over a selected span of data can be used to formulate a normalized sample statistic that is a measure of the filter's performance. Ideally, a filter operating nominally as compared to one with a suboptimal model produces a residual sequence with a smaller mean value and variance. Thus, better filter models yield smaller values of the sample statistic. The statistic is used by the GA as a measure of the fitness of an individual. Additional details on the GA as used for navigation can be found in Ref. [3].

* *"More fit" should be interpreted in a general sense as the maximization or minimization of a fitness function. In our problem, minimization was a more natural choice than maximization because the goal is to make the residuals as small in magnitude as possible.*

Illustrative Example: Stardust Small Forces Modeling

Stardust is a Discovery mission spacecraft that will fly through the cloud of dust that surrounds the nucleus of the comet, Wild 2, and, for the first time ever, bring cometary material back to Earth. A persistent Stardust navigation issue is small forces modeling. The Small Forces File (SFF) supplied by the spacecraft contractor, Lockheed Martin Astronautics (LMA), contains the operational data used for thrust impulse modeling. However, this data is prone to error. The operational trajectory determination process attempts to deal with impulse modeling errors by estimating aggregate scale factors (applied to all the impulses) driven by process noise. This approach is indirect and does not yield an impulse model that can be used confidently for prediction. A direct approach is to determine a model that operates on individual impulses. It would be prohibitive to modify the estimator (requiring significant code change) to find these impulse magnitudes because of the need for partial derivatives that relate variations of the impulses to variations of the estimated states. However, using a rapidly developed script of a candidate impulse model, the GA can efficiently search through the model's parameter space to determine appropriate values. For this example, the GA employs Stardust Doppler tracking data and the operational navigation filter. The selected impulse model employs an exponential decay between hot and cold impulses. The scale factors on the impulses and the time constant for the decay make up the GA's search space.

The script-based small forces model, coupled with the GA-based adaptive navigation tool, determined a model that yields four times better performance than results obtained using the LMA SFF. To arrive at the solution, the GA processed a week of Stardust tracking that included 1174 two-way Doppler measurements and 70 two-way Range measurements taken from January 24 through January 29, 2000. The navigation filter was setup to only estimate position and velocity components of the spacecraft. Because Stardust was experiencing no other sig-

nificant mismodeled forces or measurement effects (other than the small forces), this filter approach maximizes the information content in the residuals for the GA to extract model information. In a heuristic sense, the GA operates as a nonlinear filter to learn the appropriate small force impulse model. The January residuals using the best GA small force model and using the LMA SFF are illustrated in the upper two plots in Figure 1. The GA-based results (upper right) are 75% better than LMA SFF results (upper left). Now examine the residuals shown in the two lower plots. Both plots are residuals for data collected from March 29 to May 24. The lower left figure uses the LMA-supplied small forces file for the March to May period. The lower-right figure uses the GA-based small force model determined previously from the January tracking data. These residuals are approximately four times better than those using the LMA data, indicating that the GA-derived model is stable and robust over time.

Adaptive Navigation Using Hierarchical Mixture of Experts for Real-Time Processing Applications

Detecting and adapting filter models in real time (or near-real time) provides a robust filtering architecture that is responsive to a dynamic and changing environment. Entry/descent/landing (EDL) is a mission phase that can be characterized as highly dynamic, short lived, and having significant atmospheric model uncertainties. Autonomy and adaptability are of fundamental importance to mission success, especially for Generation 3 landing systems that seek to minimize landing error uncertainties to levels < 100 m. A promising technique, borrowed from neural networks, is called a Hierarchical Mixture of Experts (HME) algorithm.

The HME is a multilevel network of experts. Each expert has a different internal model and operates in parallel with the other experts. In our problem, the experts are individual navigation filters. A gating network continuously monitors these experts in real time and learns (via a probabilistic argument that weights the filter's residual sequence)

which filter model best matches received measurement data. In a properly configured HME, one filter has been tuned to a perceived nominal model (it acts as a control bank). The filters in the other banks have different models based on anticipated anomalies or likely modeling errors that the nominal might possess. For instance, in cruise phase common navigation error sources include solar radiation pressure mismodeling, data biases, and anomalous thrusting/outgassing events. An HME will have, at the lowest level, filter banks with model variations grouped according to the preceding classes of errors. That is, one bank of filters could have various solar radiation pressure models, another has various thrust models, and so on. Each low-level bank has a gating network associated with it that regulates the activation weight given to each filter within the bank. At the top level, another gating network regulates the weights of the low-level banks and the nominal filter. An example HME is illustrated in Figure 3. The boxes labeled with different models represent individual filters. The output of each filter is a residual sequence labeled r. The filters are grouped according to model type (i.e., SRP, etc.) that represents a bank and the output of each bank is a vector of activation weights (subscripted $g)$ that are then input into the top-level gating network that outputs activation weight vector g for the entire network.

The goal of the HME is to identify a model change, isolate its cause, and continue operations with the newly selected filter that best matches the measurement data. The HME is very robust at detecting model changes or anomalies (i.e., the residual sequences change their character from having a white noise characteristic to having a signature). Our simulations have shown that use of multiple levels in the HME yields a robust mechanism for isolating the cause of a model change (especially when considering a varied set of causes). Finally, the action of the gating network is to autonomously select the best performing filter; hence, within its designed model selection space, it resolves the problem.

Figure 1. Residuals for the January data set using the LMA-supplied SFF (upper-left plot), and the GA determined small forces model (upper-right plot). March to May residuals using the LMA SFF supplied for that period (lower-left plot), and the March to May residuals using the small forces model determined previously by the GA using the January data (lower-right plot).
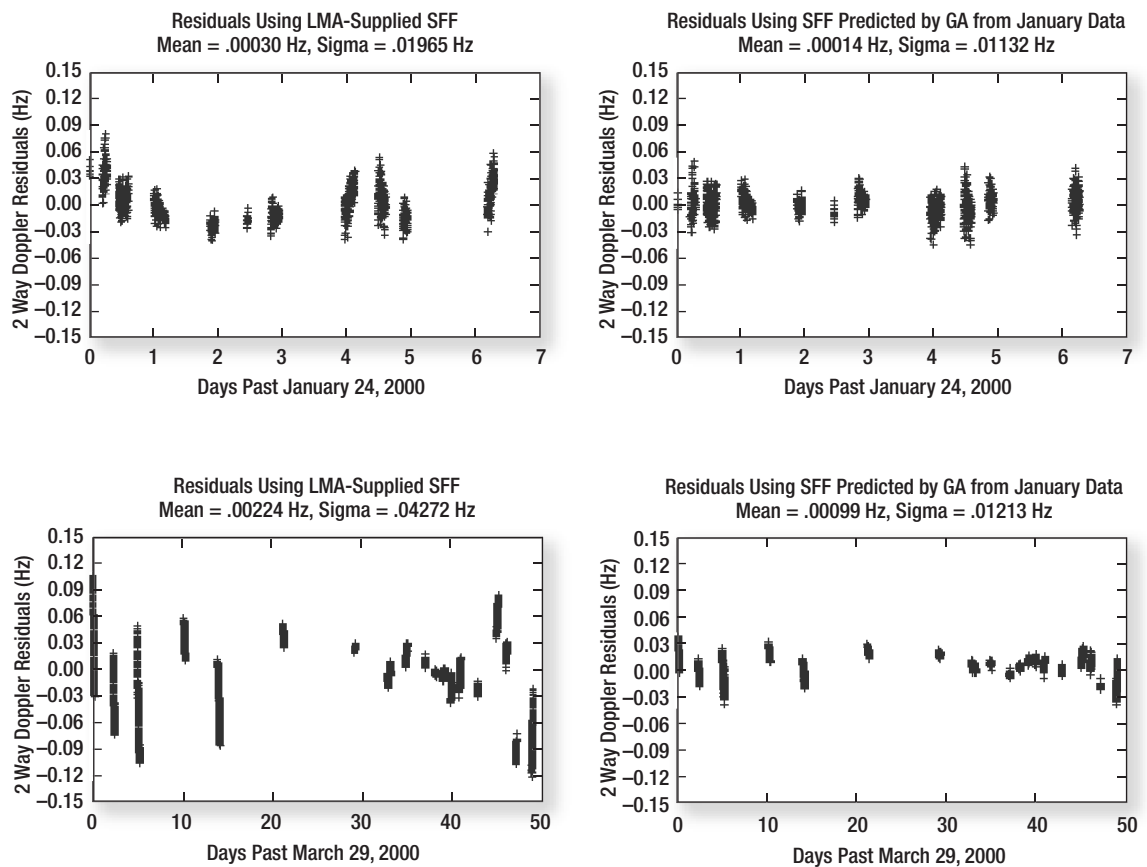
Illustrative Example: Mars Entry, Descent, and Landing

We have applied the HME technique to a cruise problem [4] processing archived operational Mars Pathfinder data in banks of filters using the Automated Real-Time Spacecraft Navigation System (ARTSN) [5]. Most recently, we have applied the method to a simulated EDL problem [6]. For EDL, our initial investigations have focused on a simplified problem using a single-level filter bank (SME) where each filter has a different representation of the atmosphere. Mars atmosphere modeling has significant uncertainties: it is not unusual to see 100% variations from prediction. It is for this reason that the filter bank is well suited to EDL, because it enables the navigation filter to adapt to the environment that is being sensed by the measurements. The experts in the SME are Extended Kalman Filters that process accelerometer data (and eventually other on-board instruments, such as gyros, altimeters, and pressure sensors) as measurements. It should be noted that processing accelerometer data in this way is new; typically, landing systems use the data in a "dead reckoning" process that directly integrates accelerations to get position (which, by implication, directly maps accelerometer errors into the solution, as well).

Each filter has a different atmospheric density profile. In real time, the gating network chooses the filter that is most likely to correctly represent the current density. To illustrate how the process might work, a filter bank is constructed using three filters with different simplified models of the Martian atmospheric density. One has an atmospheric model that is a close fit to a realistic, high-fidelity density profile model. The remaining two filter realizations have parameter values scaled by 10% during the period 250 seconds (s) to 300 s for the second filter, and by 10% during the period 300 s to 350 s, for the third filter. On receipt of a measurement, the gating network assigns an activation weight to each filter based on a probabilistic description of the measurement residual sequence (recall that the residual sequence is also used in the
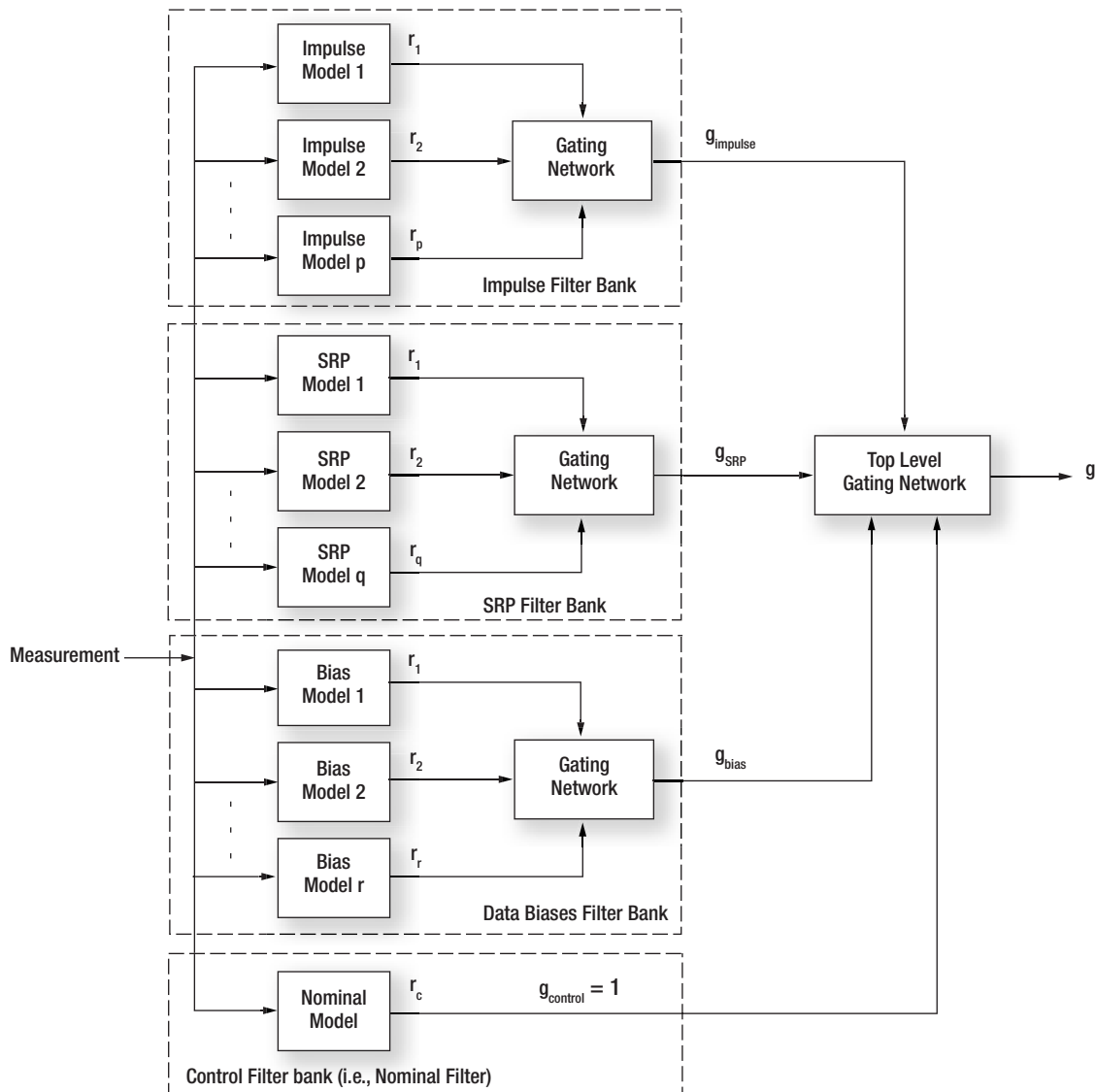
Figure 2. Illustration of a two-level HME where the low-level filter banks are separated by different models for thruster impulses, solar radiation pressure, and measurement data biases

GA application). The weights represent the probability that a given filter is the best match to the incoming data.

The gating network activation weights for the simulation are plotted in Figure 5. Initially, the three filters have the same weights (that is, each is equally probable as being the best filter). This is expected since all three have the same density coefficients. At the 250 s mark, the activation weight of the second filter begins to decline, indicating that its model does not reflect the current environment. Since we control the simulation parameters, we know that this is indeed the situation; hence, the gating network is making the correct decision: that is, filters 1 and 3 are more likely to have a better density model than filter 2. At the 300 s mark, the activation weight of the third filter begins to decline, indicating that its model representation does not reflect the current environment. Again, the gating network is making the correct decision. In the end, the first filter ends up being selected by the network as the most likely model.

Conclusions

Our work to date has focused on demonstrating the feasibility of genetic algorithms and hierarchical mixture of experts in adaptive navigation for a variety of problem do-

mains. The results have been promising. In fact, several missions have expressed an interest in using the methods to aid in their navigation tasks. Indeed, we are discussing the potential use of the GA adaptive tool in assisting navigators on Odyssey in their operational navigation mission and MER for covariance analyses. The team responsible for reconstructing the MER EDL trajectory has also expressed an interest in the techniques we have developed. Our continuing work is focused on improving the fidelity and structure of the HME for the EDL problem and in infusing the GA adaptive tool into the institutional software.

Figure 3. Activation weights for identifying the best-fit density model
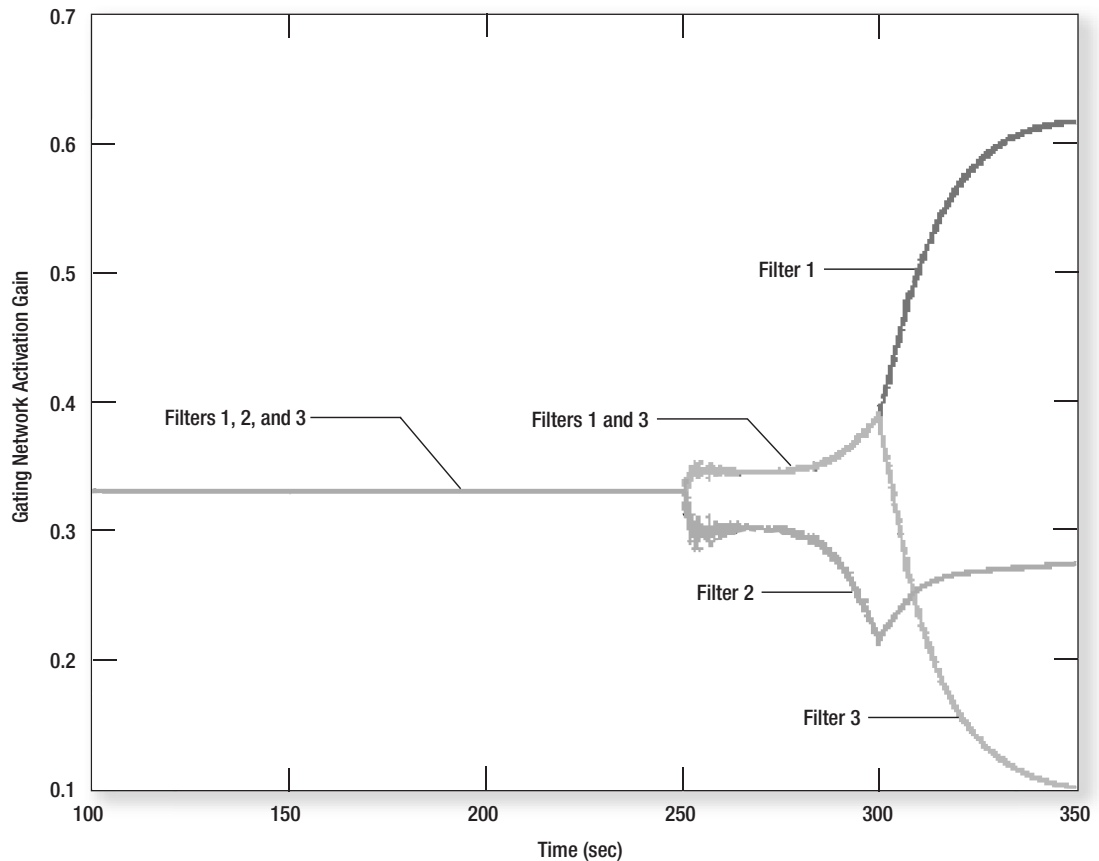
References

[1] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

[2] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, Vol. 1, Academic Press, 1979.

[3] T. A. Ely, R. H. Bishop, and T. P. Crain, "Adaptive Interplanetary Navigation using Genetic Algorithms," *Journal of the Astronautical Sciences*, Vol. 48, Nos. 2 and 3, 2000.

[4] T. P. Crain, R. H. Bishop, and T. A. Ely, "Event Detection and Identification During Autonomous Interplanetary Navigation," *AIAA Journal of Guidance, Control, and Dynamics*, accepted, to appear.

[5] L. A. Cangahaula, "ARTSN: An Automated Real-Time Spacecraft Navigation System," *TMOD Technology and Science Program News*, Issue 9, April 1998.

[6] R. H. Bishop, O. Dubois-Matra, and T. A. Ely, "Robust Entry Navigation Using Hierarchical Filter Architectures Regulated with Gating Networks," 16th International Symposium on Space Flight Dynamics, Pasadena, CA, 3–7 December, 2001.

Evolvable Hardware

DSS 13

EVOLUTION

Adrian Stoica
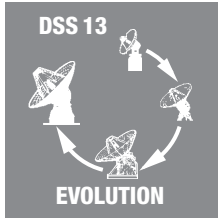
Introduction

Imagine a scenario of the future: a spacecraft that evolves in order to survive for long durations in harsh and unknown environments. The spacecraft adapts its sensors to the conditions it is sensing, modifies its antenna configuration to exploit particular orientations of its solar panels, reconfigures its electronics to provide optimal signal processing at lowest power, adapted to incoming signals and temperature, and recovers from faults. Furthermore, it learns from experience as it operates and improves its performance over time.

How far away into the future is this? Some of us believe that as early as 2010, parts of the spacecraft — probably the reconfigurable electronics and sensing — could be made evolvable. In fact, all the aspects of evolution described, admittedly for simple lab demonstrations only, have been shown possible by JPL researchers in evolvable hardware (EHW).

Design and Adaptation by Evolution

In common language, the term evolution simply denotes a capability of improvement over time. For biologists, the meaning relates to how different forms of life derive from each other and how their genetic code combines and leads to features that allow individuals to adapt and survive. Computer scientists have considered the term inspirational and crafted several types of evolutionary algorithms (EA): that is, search/optimization algorithms using representations and operators patterned after mechanisms of natural evolution, with various flavors, such as genetic algorithms, evolutionary programming, evolution strategies, etc. (Program managers are also slowly getting used to the term EA, at

first rejecting a technology that by name is merely evolutionary and not revolutionary.)

At JPL, we use the term with ambivalent semantics. We plan to build EHW that adapts/improves itself over time. On the other hand (at least for now), EA are the only techniques that have proven successful in achieving this capability; in time, other techniques may complement/replace EA. If solutions to design problems are found by EA, we refer to them as evolutionary design processes.

The following processes refer to the evolution of circuits. These are obtained through evaluations made directly on physical electronic hardware. Nevertheless, a similar process is applicable to other forms of hardware (as was demonstrated in laboratory experiments [1]). They are also applicable as to the evolution of simulations using software models.

In the context of electronic design/synthesis of an electronic circuit on a reconfigurable device, a population of candidate solutions (that is, architectural configurations) is encoded in "chromosomes." These chromosomes define the state of the switches connecting elements in the reconfigurable hardware. For example, the reconfigurable device could use cells with various resources such as operational amplifiers (opamps), logical gates, or simply transistors interconnected by switches. The chromosome determining a circuit consists of a sequence of 1s and 0s, where each value of a bit determines if the corresponding switch is closed or open.

The main steps in evolutionary synthesis of electronic circuits are explained in connection to Figure 1. First, a population of chromosomes is randomly generated to represent a pool of circuit architectures. The

chromosomes are converted into control bit-strings downloaded to programmable hardware. Circuit responses are compared against specifications/requirements (for example, comparing a target response or simply evaluating some measures of goodness); individual chromosomes are then ranked based on how well they meet the requirements. Preparation for a new iteration loop involves the generation of a new population of individuals from the pool of the best in the previous generation. Some individuals are taken as they were and some are modified by genetic operators, such as chromosome crossover (swapping fragments of chromosomial code) and mutation (random flipping of some of the bits). The process is repeated for a number of generations, resulting in increasingly better individuals. The process is usually ended after a given number of generations or when the closeness to the target response has been reached. In practice, one or several solutions may be found among the individuals of the last generation.

In principle, evolution could be performed in software simulations and then downloaded into hardware. At JPL, we have used a 128-node supercomputer running the Simulation Program Integrated Circuits Especially

(SPICE) circuit simulation program on each of its nodes, thus obtaining a parallel evaluation of the circuit models, expressed as SPICE netlists. The evolution in simulations can have advantages when experimenting with various platforms for reconfigurable devices before manufacturing the evolution/reconfiguration-oriented chips. It is also advantageous if we want to have quick tests for conditions that would take longer or be more expensive to test (for example, behavior at certain temperatures or in a situation when the time constant of the circuit to be evaluated is long). We would also need to wait seconds to get the response of the real device from hardware.

On the other hand, there are big limitations for software evolution. First, the inherent differences between models and physical hardware may cause circuits obtained from simulations to behave unsatisfactorily when tested in hardware. Such differences may be due to simplifications of the models for faster evaluations or simply because of aspects of the fabrication process. Evaluating directly in hardware is the only guaranteed means to know it will behave as such in hardware.

Second, simulations with SPICE can be lengthy and (worse) scale badly with the

Figure 1. Evolutionary synthesis/design of electronic circuits
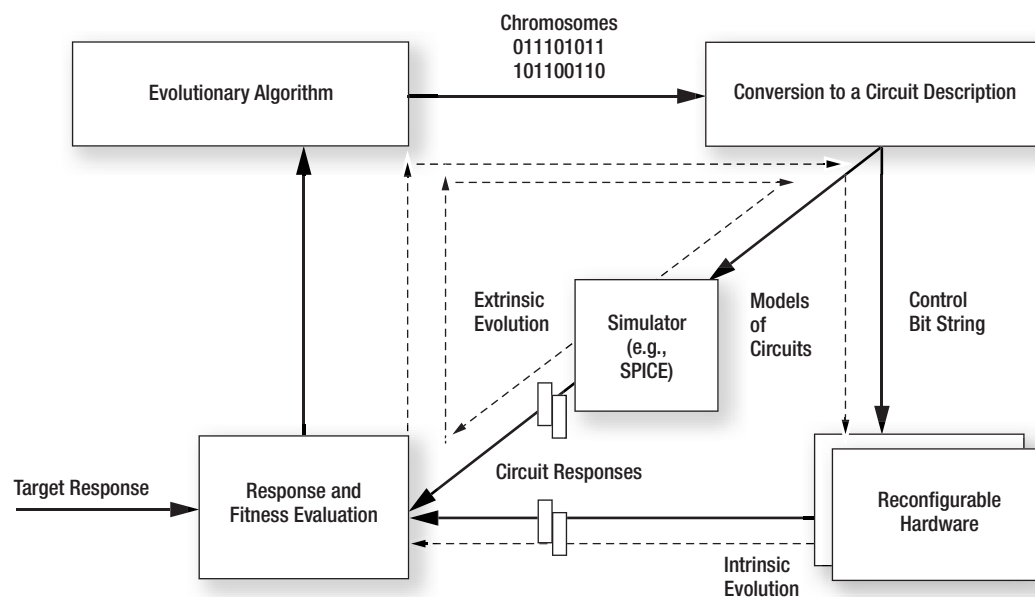
number of nodes in the circuit. Thus, although evolution of circuits with roughly 40 components currently requires tens of hours on parallel machines for millions of evaluations during one evolutionary run, it may be doable in three minutes on a desktop in 2010 [1]. Nevertheless, it is difficult to see how to scale to larger circuits using current simulators (keeping the detail; that is, without a hierarchical approach). On the other hand, evaluation directly in hardware is fast: our own JPL-designed chips evolve such circuits in seconds. Moreover, they scale perfectly: the time for obtaining a response is very much the same regardless of whether the function is implemented with 4 or 4 million transistors.

At JPL, we designed and built three generations of "evolution-oriented" chips, reconfigurable at the transistor level. These are referred to as field programmable transistor arrays (FPTA). The latest chip, with a 5×7 square millimeter (mm^2) die in Taiwan Semiconductor Manufacturing Company's (TSMC) 0.18 micron process, is designed as a 8×8 matrix of reconfigurable cells. It can receive 96 analog/digital inputs and provide 64 analog/digital outputs. Each cell is programmed through a 16-bit data bus and requires 5000 bits to program the whole chip. It may be the world's first true field programmable mixed-signal array (FPMA), in the sense that its transistor-level reconfiguration allows mapping both digital and analog circuits. The chip can configure 64 opamps, more than the largest commercial field programmable analog array (FPAA) chip, which has only 20 opamps (the opamp equivalent performance is poorer, however). Each cell also has reconfigurable resistances and capacitors.

Evolved Circuits and Recovery from Faults

A variety of traditional and non-traditional circuits were evolved on FPTA chips, as well as in simulations using the FPTA architecture and in unconstrained evolution where designs did not need to fit a programmable cell model. Traditional functional circuits include filters, operational amplifiers, logical gates, D/A converters, rectifiers, analog multipliers, functional approximators for fuzzy controllers, analog computational circuits such as gaussian circuits, etc.

Polymorphic circuits are an example of nontraditional designs obtained by evolution. Polymorphic circuits refer to electronics with superimposed, built-in functionality [3]. A function change does not require switches/reconfiguration as in traditional approaches. Instead, the change comes from modifications of operational points. These move to different characteristic curves of the devices involved in the circuit in response to controls such as temperature, power supply voltage (V_{DD}), control signals, light, etc. For example, a temperature-controlled polytronic AND/OR gate behaves as AND at 27 degrees Celsius (° C) and as OR at 125° C. We successfully evolved polytronic circuits in which the control is done by temperature, morphing signals, and V_{DD}, respectively. Figure 2 shows such a circuit controlled by V_{DD}. Because of their capacity of storing/hiding "extra" functions, such circuits have an intrinsic provision for a form of watermarking or conceiving extra functionality, thus finding possible uses in intelligence/security applications.

In a normal evolutionary run, the fitness of the best circuits improves over generations until a certain performance level (hopefully satisfactory) is reached. The fact that the search is carried by a population of solutions in which many solutions may differ only by a few bits provides interesting fault-tolerant characteristics. For example (Figure 3), in experiments in which we injected faults, such as circuit interrupts, restarting evolution with the last population obtained in evolving the solution that was damaged provided much faster (about a third of the time) recovery (here to about 90% of original performance) than starting evolution with random candidates, which took approximately the same time as it took evolving the first solution.

Another area where evolution has proven successful is recovery of circuits with functionality degraded at extreme temperatures. Experiments with bulk silicon at −196° C and +320° C show recovery of functionality by redesigning the circuit on-chip, obtaining a new configuration that can provide the desired response at that temperature. Supple-

Figure 2.
(a) An evolved polymorphic circuit controlled by supply voltages.
(b) Inputs and output transient response for V_{DD} = 1.2 V (left) and V_{DD} = 3.3 V (right)
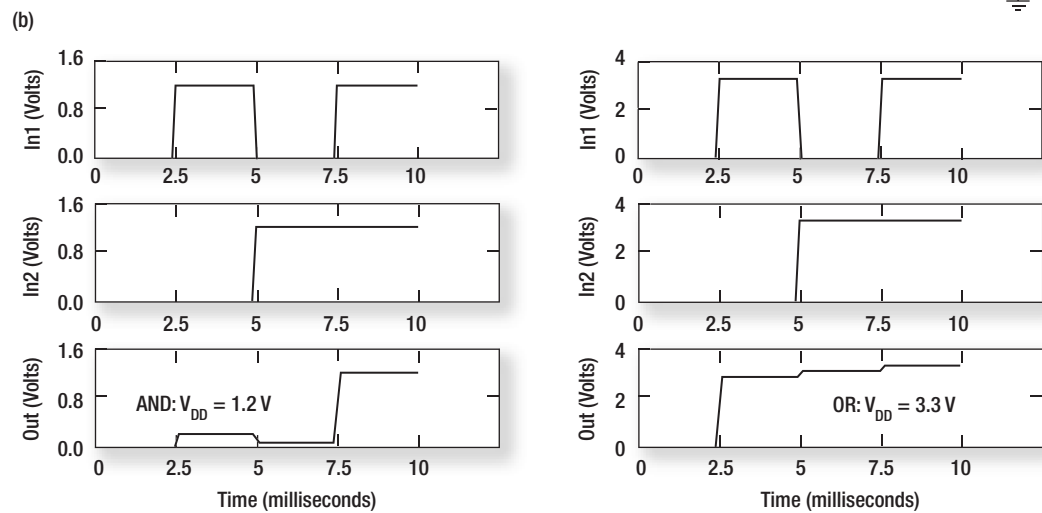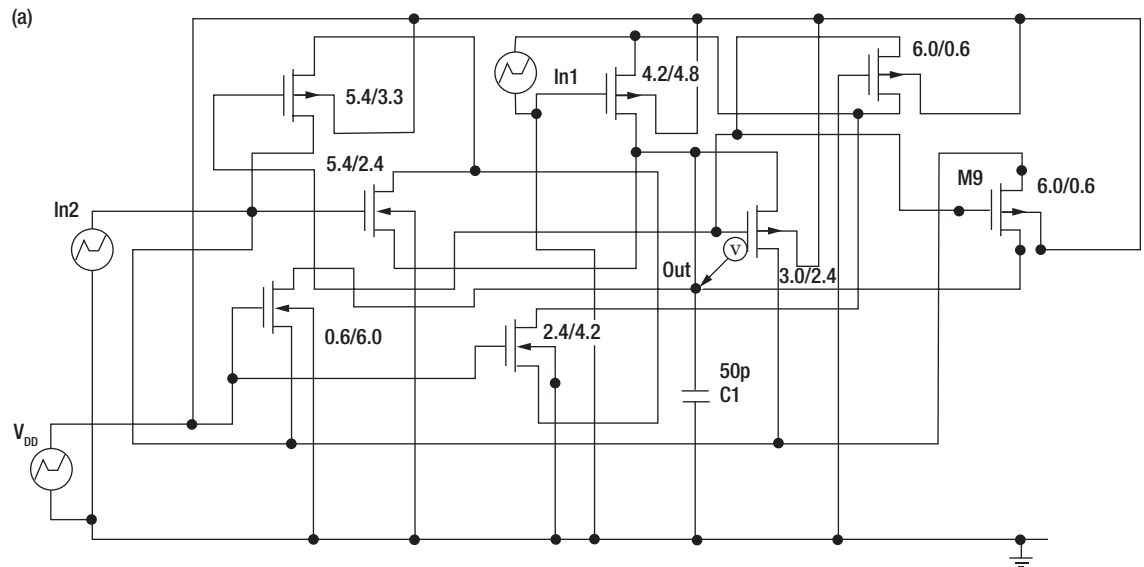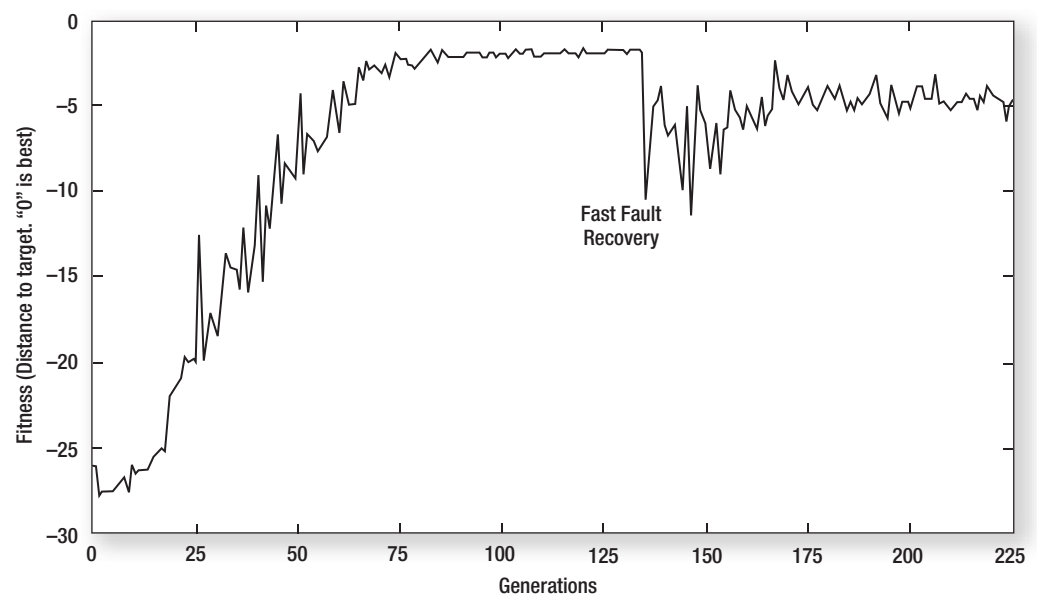

Figure 3. Improvement of fitness value characterizing the performance of the circuit

menting device/material solutions, circuit solutions provided by EHW could further expand the operational range of extreme electronics.

Limitations, Challenges, and Work in Progress

Design by evolution is a "what you ask is what you get" process. The bad news is that, unless you ask carefully, evolution has many ways of tricking and finding easy (and not what we really want) solutions to requirements/specs. For example, say we ask for a logical gate by giving a logical table and assume certain voltage levels for high and low. Suppose we perform evolution using a transient analysis at the order of nanoseconds and, after a number of generations, we obtain a circuit that satisfies the requirements. Now we test on a different timescale — milliseconds — and we find that the response is wrong: the solution that was obtained only works at a certain timescale! Alternately, if we evolve at a milliseconds timescale, the resulting circuits may be slow and will not work at nanoseconds. Alternately, if we test at a temperature different from the one we used in evolution, the solution circuit may also not operate properly. A possible way to address this is to include several conditions in evaluation and give combined fitness functions. However, do we have to explicitly specify behavior in all possible contexts? If so, at what resolution? Can we predict or guarantee behavior in a region that was not evaluated during evolution? Evolution exploits a large space of solutions and has to pick some; the specifications help this narrowing search process, yet sufficient specification needs to be made. Sometimes we fail at this since, in common designs, we use some implicit assumptions that we take for granted. This does not work in evolutionary design.

Evolvable hardware is not ready yet for space, but it is getting closer. And, in a not too distant future, it will bring two main benefits to spacecraft survivability: it will help preserve functionality in conditions where hardware is affected by faults, aging, temperature drifts, radiation, etc. and will allow *in situ* automatic generation of new designs for new functions when required (for example, by unexpected opportunities or mission changes).

About the Evolvable Hardware Research at JPL

Adrian Stoica initiated JPL's EHW research in 1996 and has since then led the effort as part of Anil Thakoor's Bio-inspired Technology and Systems Group in Section 344. With Anil's support, Adrian obtained initial funding through Satish Khanna, JPL Chief Scientist Mous Chahine, and David Collins. A year later, Evolvable Hardware technology was selected by Leon Alkalai for the first round of funding of the new Center for Integrated Space Microsystems and, in 1998, a $2 million dollar contract from the Defense Advanced Research Projects Agency's (DARPA) Adaptive Computing Program was won. Currently, the research is funded jointly by DARPA and NASA, mainly through programs managed by Benny Toomarian and Neville Marzwell. Initially a one-man effort, the team has grown by attracting other leading researchers in the field, such as Didier Keymeulen from Japan's Electrotechnical Laboratory and Ricardo Zebulum from University of Sussex, both centers of excellence in the field. The team, which also includes two Ph.D. students and other collaborators working part-time, is now the strongest team in the world, measured by the number of papers published, representation in program committees of international conferences, editorial boards of dedicated journals, etc. The team is also the catalyst in focusing the EHW community to addressing problems of interest to NASA and DOD and, since 1999, has been organizing the only annual international meeting dedicated to the EHW field. The 2002 NASA/DOD Conference on Evolvable Hardware will be hosted by NASA Goddard, July 15 through 18, 2002.

References

[1] http://cism.jpl.nasa.gov/ehw/

[2] J. R. Koza, F. H. Bennett III, D. Andre, and M. A. Keane, "Genetic Programming III," *Darwinian Invention and Problem Solving*, Morgan Kaufman, San Francisco, 1999.

[3] A. Stoica, R. Zebulum, and D. Keymeulen, "Polymorphic Electronics," in *Proceedings of the International Conference on Evolvable Systems*, Springer, 2001.

For more information, papers, and video clips showing evolution of various circuits, visit our website at http://cism.jpl.nasa.gov/ehw
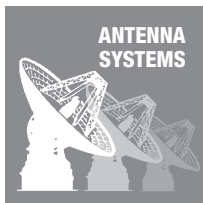
Raster-Scan for Calibration of DSN Antennas

ANTENNA
SYSTEMS

*David J.
Rochblatt,*

Paul H. Richter,

Michael J. Klein,

and

Phil Withington

Introduction

NASA's Deep Space Network (DSN) of large radio antennas provides the basic communications links with the many spacecraft that form part of the nation's unmanned space exploration program. In order to satisfactorily carry out this mission, each antenna must undergo various calibrations to ensure that it is operating in the most efficient manner possible and, hence, delivering maximum information at minimum cost.

As part of the strategy of improving the overall performance capability of the DSN, there has been a steady increase in the operating frequency of these antennas over the years, going from S-band ($f \approx 2.3$ gigahertz [GHz]), to X-band ($f \approx 8.4$ GHz), and, most recently, to Ka-band ($f \approx 32$ GHz). Assessment of the performance of the R&D 34-meter (m) beam waveguide (BWG) antenna DSS 13 at W-band ($f \approx 90$ GHz) is underway.

One can gain a better appreciation of the implications of these frequency increases for antenna calibration by considering the corresponding wavelengths of the radiation. Thus, for S-band $\lambda \approx 13$ centimeters (cm), for X-band $\lambda \approx 3.6$ cm, and for Ka-band $\lambda \approx 0.9$ cm. The essential performance characteristics of an antenna, such as pointing capability and aperture efficiency, are strongly dependent on the wavelength of the radiation being detected. That is why the large (14:1) decrease in wavelength has resulted in the need for much greater precision of such parameters as reflector surface figure, azimuth track smoothness, subreflector and BWG mirror alignments, etc.

Calibration procedures require accurate measurement of the various parameters of interest. In the case of antenna calibration, these fall naturally into two categories: those derivable from the measurement of amplitude and phase of a received coherent microwave signal and those derivable from the measurement of received power from a noncoherent source, such as a radio star.* The former case involves the interference of received signals from the antenna under test (AUT) and a small, reference antenna mounted nearby, using a technique known as *microwave holography* [1]; the latter utilizes a total power radiometer (TPR) measuring system to determine the antenna temperature of the source.

In both cases, it is necessary to mechanically scan the antenna beam past the source in some manner; the way this is accomplished turns out to be crucial to the success of the measurement.

As a result of a systematic analysis of the entire measurement procedure, with particular attention to the noise characteristics of the TPR, plus the tropospherically induced radiometer fluctuations and the implementation of new techniques for data acquisition and reduction, it is possible to obtain measurement precision yielding up to an order of magnitude improvement over previous methods in the determination of antenna aperture efficiency, and factors of five or more in the determination of pointing errors and antenna beamwidth. This improvement has been achieved by performing *on-the-fly* (OTF) mapping of both extended and point radio sources [2, 3].

The significance of improvements in antenna calibration, performance, or performance evaluation in the DSN can be put in perspective by recognizing that each decibel (dB) of improvement in the quantity gain-to-noise temperature (G/T) is estimated to be

---

* *The term "radio star" refers to any natural, compact source of radiation, which, for example, may actually be a remote galaxy of stars, a region of bright microwave emission in our own Milky Way galaxy, or a planet in our solar system.*

worth about $6.8 M / % ($160M / dB) in terms of mission support capability [R. Clauss, private communication].

The advantages of the OTF mapping technique over the traditional, or boresight, approach include the following:

1. Removes a major source of error in determining antenna efficiency by eliminating the need for independently derived source-size correction factors.

2. In determining the optimum integration time during continuous scan measurements, the actual spectrum of the total power radiometer noise fluctuations is taken into account. This includes radiometer flicker noise and tropospheric turbulence effects.

3. It provides a direct comparison of the flux density of one radio source with that of another, thus enabling the accurate calibration of many radio sources for future antenna calibrations.

4. A highly accurate, repeatable, and fully automated system will provide reductions in cost and improve the reliability of antenna performance.

While other calibration techniques have advantages for certain applications, the OTF mapping accurately provides all of the required calibration data in the shortest measurement period and includes the beam shape as well as the source flux density.

In the following section we describe the scanning process in more detail and consider the merits of the so-called *raster-scan*, more generally known as the OTF mapping method, over other approaches.

Scanning a Radio Source

For the OTF mapping measurements utilizing a TPR, the desired products are the peak temperature corresponding to the source, the pointing offset when the antenna is commanded to the source, and the shape of the antenna beam main lobe. These parameters in turn permit one to calculate the antenna aperture efficiency by comparing the measured peak temperature with that expected for a 100% efficient antenna, the mechanical pointing error resulting from small

misalignments of various parts of the antenna structure, and subreflector and other mirror misalignments.

Regardless of the scan geometry chosen, it is necessary to make these measurements as a function of antenna beam elevation and azimuth so that the dependence of antenna performance on these angles may be determined and compared with that expected from theory.

For the TPR measurements, the sources must be tracked across the sky during the measurement period so that obtaining good angular resolution in elevation and azimuth, as well as good all-sky coverage for pointing model determination, requires relatively short measurement periods, thus implying relatively fast scans.

A third problem is that many natural radio sources used for antenna calibration purposes are only approximately point-like, with the result that a given antenna may not collect all of the radiation emitted by the source; i.e., the antenna may partially resolve the source structure. In such cases, one must estimate how much of the total known radiation from the source would actually be collected by the antenna under test at the frequency being used if it were 100% efficient. The resulting estimate, leading to a source-size correction factor, introduces another degree of uncertainty to the measurements, which can be avoided by using the OTF mapping approach.

The OTF Mapping Method

The key to this approach is maintaining a constant, known angular velocity of antenna motion during the taking of data along a given direction. To ensure alignment of the individual subscans within the full raster, a position-based triggering approach was taken for the design of the data acquisition. In this design, the first data point of each subscan is determined by position measurements, while the rest of the subscan data points are determined by timing, such that the TPR is sampled at a constant known rate so that the relative position at which the data are taken is known with high accuracy and antenna-settling time is no longer an issue. The accurate determination of the first data point for each

of the raster subscans was obtained by direct readout of the antenna angle encoders for the azimuth and elevation angles at a high rate (100 samples per second). The OTF mapping data acquisition algorithms include computation of the radio source (or planets) position such that, at any given time, the position of the antenna relative to the source is known. In this way, a synchronized data acquisition for antenna position and TPR data was achieved.

Since the data are taken "on the fly," the integration occurring during the sampling interval results in an attenuation of high-frequency information; however, this can be recovered by an inverse filtering process (Wiener filter).

Since the source is scanned in a raster type pattern by stepping from line to line, a complete data set corresponding to a complete raster contains all of the relevant data, not just a sampling of it along two orthogonal directions. This means that one is effectively including all of the source radiation so that no source-size correction is necessary. The resulting 2-dimensional data set may then be used to determine, by means of least-squares fitting, the best-fit main beam pattern from which the relevant calibration parameters may be directly determined. Figure 1 shows the *raster–scan TPR* system architecture.

Advantages of the OTF Mapping System for Antenna Calibration

The OTF mapping system was used at DSS 13 to support the calibration of the 13.8-GHz receiver built for the Cassini–Jupiter Microwave Observing Campaign (Cassini–JMOC) [4]. The OTF was used to evaluate the RF performance of the antenna and the receiving system and to calibrate the aperture efficiency with high accuracy. The OTF system is especially suited to detect and quantify changes in the main beam and the side-lobes of the antenna pattern as the antenna tracks radio sources over wide ranges of azimuth and elevation; this feature is being used to improve the precision of the calibration measurements for Cassini–JMOC.

To measure the gain of large antennas, one can measure the received power from a radio source that has been previously calibrated by

independent means [5, 6, 7]. A practical method to accomplish this is to measure the received power from the calibration radio source and compare the result with the theoretical result one would measure with a "perfect" antenna. Expressed as a ratio, the result is the aperture efficiency $\varepsilon(\psi)$, where ψ represents the orientation of the antenna; e.g., azimuth and elevation. The formal expression for $\varepsilon(\psi)$ is:

$$\varepsilon(\psi) = 2kT_c(\psi) / AS_c \qquad (1)$$

where T_c is the increase in the system noise temperature when the antenna is pointed precisely "on" the calibration radio source with flux density S_c (W/m^2 Hz). The antenna aperture, A, for a circular dish of diameter d is the geometrical area $A = \pi d^2/4$.

In the real world of radio astronomy, this process is complicated by the fact that sources whose radio frequency flux densities have been accurately measured tend to be rather large in angular size relative to the antenna pattern of a large-aperture radio telescope like the DSN antennas. In practice, the small angular width of the antenna beam partially resolves the angular area of the radio source with the result that some of the radio flux density is not collected by the antenna when it is pointed "on source." A "correction for source size," $C_s(\psi)$, is typically used to compensate for this effect. Note that the value of C_s is expected to vary with antenna orientation (ψ) because both the beam shape and the angular orientation of the radio source change as source rises and sets across the sky.

The OTF mapping system can eliminate for this partial resolution effect in a rather elegant manner if one measures the raster scan patterns of a calibration source and a highly compact (point-like) radio source at nearly the same elevation angle. The raster scan detects all of the flux from both sources, regardless of angular size, because the output is the two dimensional convolution of the antenna beam and the source brightness distribution. The output can be visualized as a response volume. An example is shown in Figure 3(a), which shows a raster scan of Venus taken at DSS 13 as part of the Cassini–JMOC calibrations. Venus was close to Earth at the

Figure 1.
OTF mapping
computer
interface
with the
Cassini–JMOC
radiometer

time, so its angular diameter was large enough to be partially resolved by the antenna beam; i.e., the response pattern is slightly broader than the antenna pattern itself. To complete the process, an OTF map of a point-like radio source would be made a at the same elevation (≈ 58 degrees) and the ratio of the two response volumes would equal the ratio of the two flux densities as shown in Equation 2:

$$C_s(\psi) = \frac{S_c}{S_{pt}} = \frac{T_{pt}(\theta,\phi)\iint T_c(\theta,\phi)\big|_{\psi_o} d\Omega}{T_c(\theta,\phi)\iint T_{pt}(\theta,\phi)\big|_{\psi_o} d\Omega} \qquad (2)$$

Where (θ,ϕ) are the rectangular, angular coordinates relative to the source center and $T(\theta,\phi)$ are the source temperature recorded at these observation points. The center of the map is at a particular value of ψ_o.

Rewriting Equation 1 for the point-like source with parameters Tpt and Spt and using Equation 2 to substitute for Spt (an a priori unknown value), one gets a new expression for $\varepsilon(\psi)$:

$$\varepsilon(\psi) = \frac{2kT_{pt}(\psi;\theta,\phi)\iint T_c(\theta,\phi)\big|_{\psi_o} d\Omega}{AS_c \iint T_{pt}(\theta,\phi)\big|_{\psi_o} d\Omega} \qquad (3)$$

All of the terms on the right hand side of Equation 3 are either known values or they are measured by the OTF system. This result illustrates some of the valuable features of the OTF mapping system:

(a) The need to derive source size correction factor $C_s(\psi)$ is eliminated. However, one has the option to calculate $C_s(\psi)$ for various calibration radio sources using the OTF results and Equation 2 if that is desirable.

(b) The need to derive accurate values of $T_c(\psi)$ for extended sources is eliminated. This is desirable because accurate derivations of $T_c(\psi)$ require mapping and deconvolving source structure from the antenna patterns which change with antenna orientation (ψ). Using the OTF mapping technique, the accuracy of the antenna efficiency measurement is limited by that with which the calibration source flux density S_c is known.

(c) The OTF mapping system enables the experimenter to use compact, point-like sources as secondary calibration sources for precision antenna calibrations. The vast majority of these radio sources are quasars, which are so distant that their angular sizes are very small. The problem is their radio brightness (flux density) values are highly variable; therefore, one must calibrate them against the handful of absolutely calibrated radio sources that are available. With few exceptions, the timescales of the quasar variations are typically a few days; therefore, flux density calibration measurements can be done rather infrequently and relative measurements of antenna performance with azimuth and elevation can be done almost any time.

Initial Results of Cassini–JMOC Measurements

Tests were run in April 2001 at DSS 13 using the OTF system to observe Venus, Jupiter, 3c405, 3c273, 3c274, and 3c84. Given that the antenna half power beamwidth (HPBW) for the 34-m antenna operating at 13.8 GHz is approximately 0.041 degrees, we scaled the raster dimensions for approximately three times the HPBW to produce a map 0.125 × 0.125 degrees on the sky. The temporal resolution along the scan was about 1/10 HPBW (approximately 0.004 degrees) which resulted in data arrays of 33 × 33 points. Typical raster-scans required 15 to 20 minutes to complete. When the weather was calm, excellent raster alignment was achieved. When the wind was above 10 miles per hour, apparent misalignment in the raster was observed (see Figure 5, 34.3-degrees elevation). The source of this problem is believed to be subreflector oscillations because no apparent misalignment in the antenna angle encoders registering was noticed. Additional tests are recommended to validate and study this result.

Figures 3(a) through 3(c) illustrate the stages of data processing for the OTF system. Figure 3(a) shows the 3-D response of the raster scan data taken across Venus when it was near 58 degrees elevation. Venus was close to the Earth and, therefore, it was a very strong radio source when the measurements were made. Consequently, the noise level in the map is very low and the smooth surface of
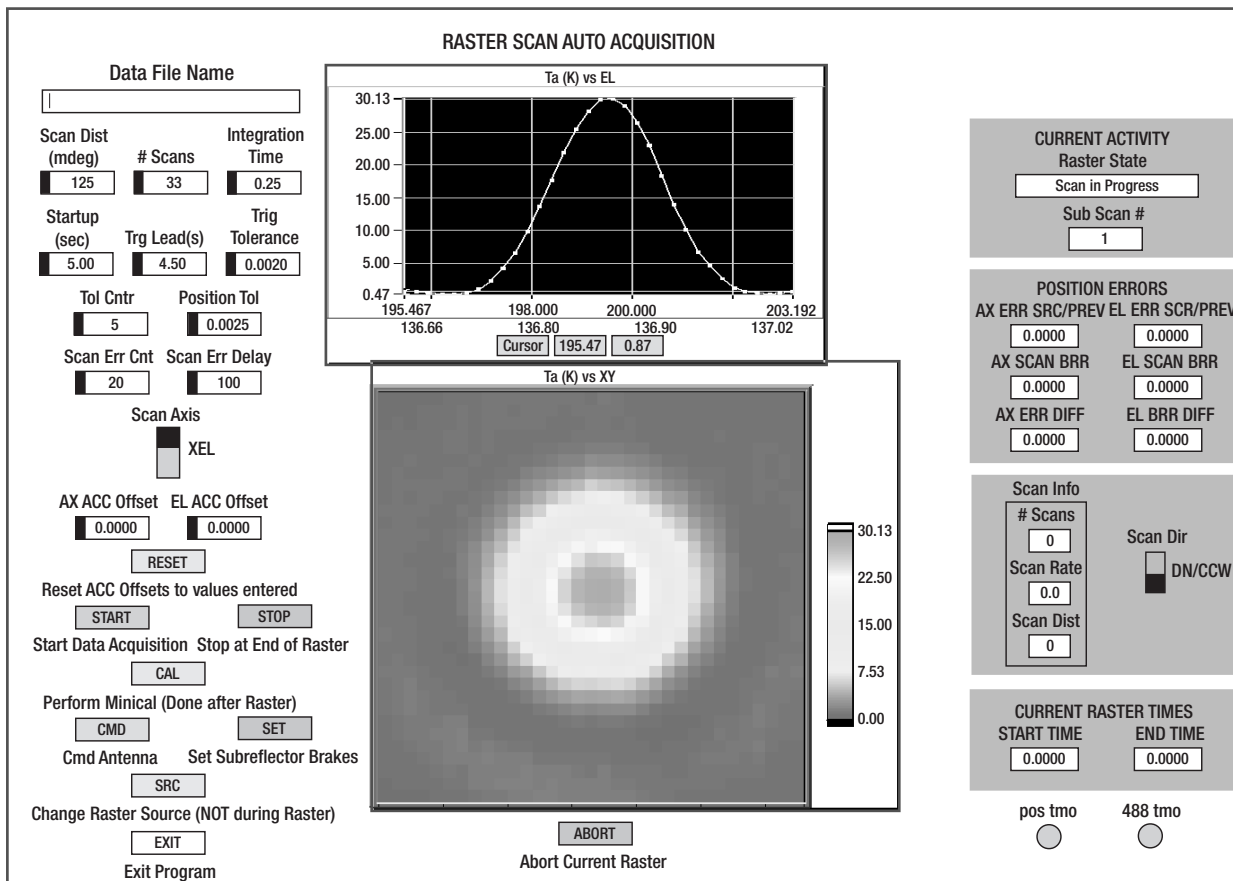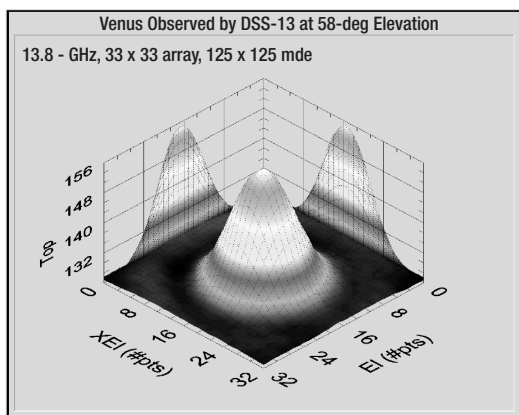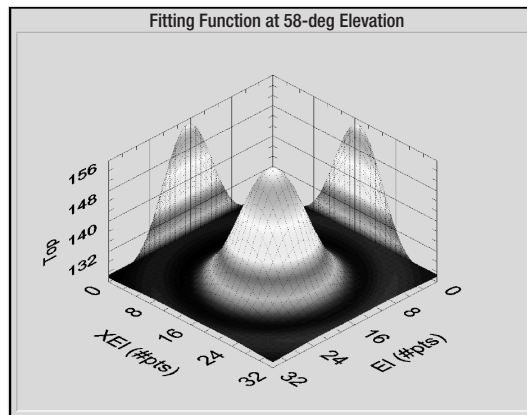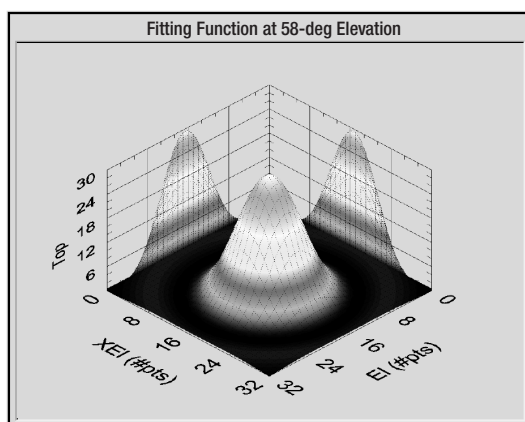
RASTER SCAN AUTO ACQUISITION

Data File Name

| |

Scan Dist (mdeg): 125
Scans: 33
Integration Time: 0.25

Startup (sec): 5.00
Trg Lead(s): 4.50
Trig Tolerance: 0.0020

Tol Cntr: 5
Position Tol: 0.0025

Scan Err Cnt: 20
Scan Err Delay: 100

Scan Axis: XEL

AX ACC Offset: 0.0000
EL ACC Offset: 0.0000

RESET
Reset ACC Offsets to values entered

START   STOP
Start Data Acquisition Stop at End of Raster

CAL
Perform Minical (Done after Raster)

CMD   SET
Cmd Antenna   Set Subreflector Brakes

SRC
Change Raster Source (NOT during Raster)

EXIT
Exit Program

Ta (K) vs EL

30.13
25.00
20.00
15.00
10.00
5.00
0.47
195.467   198.000   200.000   203.192
136.66   136.80   136.90   137.02

Cursor 195.47 0.87

Ta (K) vs XY

ABORT
Abort Current Raster

30.13
22.50
15.00
7.53
0.00

CURRENT ACTIVITY
Raster State
Scan in Progress
Sub Scan #
1

POSITION ERRORS
AX ERR SRC/PREV EL ERR SCR/PREV
0.0000   0.0000
AX SCAN BRR   EL SCAN BRR
0.0000   0.0000
AX ERR DIFF   EL BRR DIFF
0.0000   0.0000

Scan Info
Scans: 0
Scan Rate: 0.0
Scan Dist: 0
Scan Dir: DN/CCW

CURRENT RASTER TIMES
START TIME   END TIME
0.0000   0.0000

pos tmo   488 tmo

Venus Observed by DSS-13 at 58-deg Elevation

13.8 - GHz, 33 x 33 array, 125 x 125 mde

156
148
140
132
0

Tcp

XEl (#pts)
8 16 24 32
EI (#pts)
0 8 16 24 32

3(a)

Fitting Function at 58-deg Elevation

156
148
140
132
0

Tcp

XEl (#pts)
8 16 24 32
EI (#pts)
0 8 16 24 32

3(b)

Fitting Function at 58-deg Elevation

30
24
18
12
6
0

Tcp

XEl (#pts)
8 16 24 32
EI (#pts)
0 8 16 24 32

3(c)

Figure 2. OTF mapping computer display during data acquisition

Figure 3. Graphic 3-D displays illustrate the OTF processing steps and demonstrate the excellent raster alignment achieved. The observed brightness response scanning across Venus (a) is hardly distinguishable from the 3-D Airy fitting function (b).

the 3-D plot indicates the excellent alignment of the individual raster sub-scans. The x and y-axis coordinates correspond to elevation (El) and cross elevation (XEl) of the maps. The samples along sub-scan direction are approximately 0.0039 degrees. The z-axis shows the measured system noise temperature in Kelvin units.

Figure 3(b) is the result of fitting the raw data with an Airy function, which is a mathematical expression that describes the spatial smoothing caused by diffraction when radio waves (or light waves) are reflected off a circular aperture, which in this case is the 34-m parabolic dish. The formal equation for diffraction of radio antennas is the Jacobi-Bessel series expansion of the far-field pattern of the antenna [8].

Figure 3(c) is a plot of the fitting function after removing the background noise, which is fitted to a two-dimensional baseline with arbitrary slope.

Figure 4 shows the raw OTF mapping data of Venus at 8.7 degrees elevation next to the 3-D Airy function derived by non linear least squares fit. From the fitting function parameters, we compute the peak source temperature, the pointing offsets in two orthogonal axis (El and Xel), antenna beamwidth in the two axes, and the constant and slope of the background temperature. The plotting of the fitting functions on the right hand side of the plot was done after the removal of the background temperature and the computation of the functions integral followed. These plots demonstrate the capability of the OTF mapping system to reveal distortions in the antenna beam pattern with low elevation angles. A series of patterns taken of Venus at source elevations from 58 to 8.7 degrees is

displayed in Figure 5. These OTF maps clearly show the distortion of the antenna main beam at low elevation angles (8.7, 12.7, and 16.6 degrees), which are due to gravity-induced deformation of the main reflector surface [1]

In deriving the antenna efficiency curve in Figure 6, we used the OTF mapping data from 3c273 (point source) and 3c274 (Calibrated source), both obtained near the rigging angle of 49 degrees, where an efficiency of approximately 61% was computed. Then the OTF mapping data of Venus obtained from 8.7 to 58 degrees elevation was calibrated at 49 degrees elevation to that value. In the process, a source size correction of 1.070 was computed for 3c274 near the same rigging angle, which compares well with the independently derived value of 1.075 obtained using the other scanning techniques requiring many more observations.

This application of the OTF system for the Cassini–JMOC calibration work demonstrates some of the valuable attributes of the OTF system for the calibration and performance analysis of DSN antennas. Future work will extend the raster-scan analysis of strong sources to include the study of sidelobes in the antenna far-field beam. It is anticipated that measurements with sufficient precision will be made to enable the application of phase retrieval techniques for the derivation of antenna surface deformation over elevation angles suitable for the calibration of a deformable flat plate [9, 10].

References

[1] D. J. Rochblatt, "Holographic Measurements of the NASA-JPL Deep Space Net-

Figure 4. A raster of Venus at 8.7 degrees elevation (left panel) and the fitting function after the removal of the background noise (right panel)

El: 8.7-deg   12.7-deg   16.6-deg   34.3-deg   44.6-deg   54.2-deg   58-deg
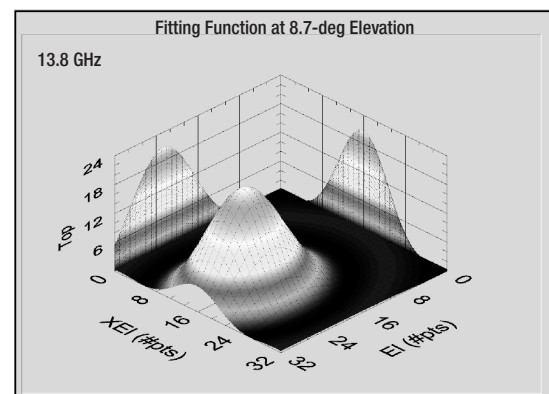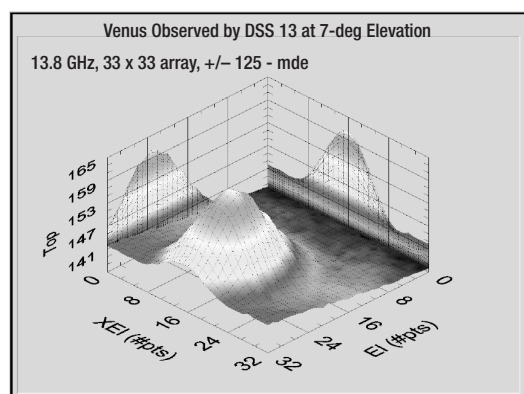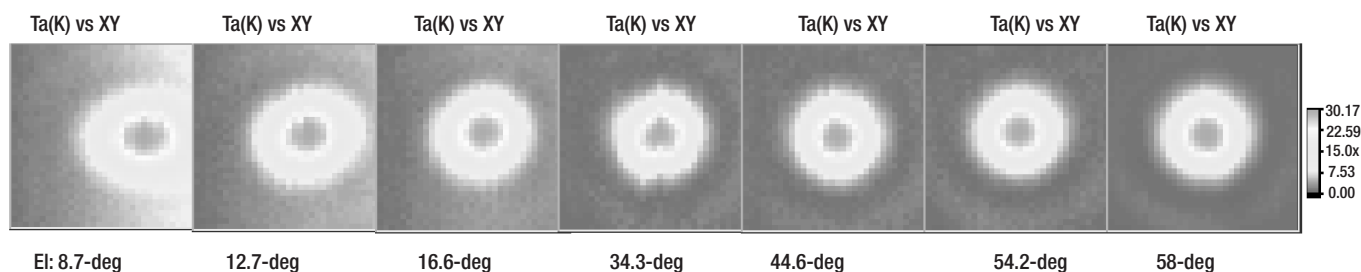
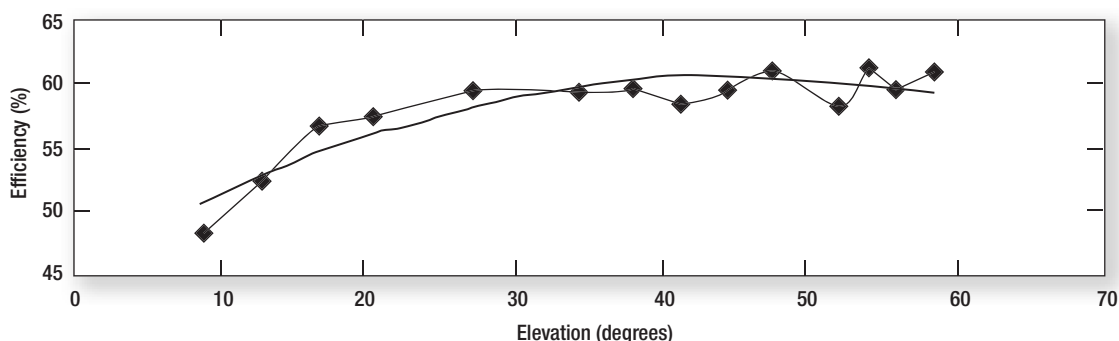Figure 5. Patterns of Venus at source elevations from 58–8.7 degrees

Figure 6. DSS-13 efficiency derived by OTF mapping technique

work Antennas," *Proceedings IEEE Aerospace Conference*, Snowmass at Aspen, Colorado, March 21–28, 1998, on CD.

[2] P. H. Richter and D. J. Rochblatt, "Raster-Scan for Calibration of DSN Antennas," *TMOD Technology and Science Program News*, Issue 9, April 1998.

[3] P. H. Richter and D. J. Rochblatt, "A Microwave Performance Calibration System for NASA's Deep Space Network Antennas, Part 1: Assessment of Antenna Gain and Pointing, and Calibration of Radio Sources," *Proceedings, 10th Int. Conf. on Antennas and Propagation*, pp. 1.142–1.149, Vol. 1, 14–17 April 1997.

[4] M. J. Klein, S. J. Bolton, A. J. Freiley, S. Gulkis, M. A. Janssen, S. J. Levin, D. J. Rochblatt, J. P. Roller, and R. K. McLeod, "DSN and GAVRT Observations of Jupiter at 13 GHz and the Calibration of the Cassini Radar Instrument for Passive Radiometer," URSI Symposium, 2002.

[5] L. W. M. Baars, R. Genzel, I. I. K. Pauliny-Toth, and A. Witzel, "The Absolute Spectrum of CasA: An Accurate Flux Density Scale and a Set of Secondary Calibrators," *Astron. Astrophys.* 61, 99–106, 1977.

[6] M. J. Klein and C. T. Stelzried, "Calibration Radio Sources for Radio Astronomy: Precision Flux Density Measurements at 2295 MHz," *Astron. J.* 81, No. 12, 1078–1083, 1976.

[7] P. H. Richter and S. D. Slobin, "DSN 70-Meter Antenna X- and S-Band Calibration Part I: Gain Measurements," *TDA Progress Report* 42-97, 315-351, Jan–Mar, 1989.

[8] V. Galindo-Israel and R. Mittra, "A New Series Representation for the Radiation Integral with Application to Reflector Antennas," *IEEE Trans. Antennas Propagat.*, Vol. AP-26, p. 628, July 1978.

[9] D. J. Rochblatt and V. Vilnrotter, "Demonstration of a Ka-Band Array Feed–Deformable Flat Plate Compensation System on the 70-m Antenna," *JPL Technical Science Newsletter #12*, May 2000.

[10] D. J. Rochblatt, Daniel Hoppe, William Imbriale, Manuel Franco, Paul Richter, Phil Withington, Herschel Jackson, "A Methodology for the Open Loop Calibration of a Deformable Flat Plate on a 70-meter Antenna," *Proceedings of the Millennium Conference on Antennas & Propagation* AP2000, Davos, Switzerland, 9–14 April 2000.

IND Technology Program Office
Jet Propulsion Laboratory
MS 303-407
4800 Oak Grove Drive
Pasadena, California 91109-8099